Technical Note

1969-47

J. E. Evans

Probability Density Function Estimation (with Applications to Receiver Design for Reception in Non-Gaussian Noise)

29 August 1969

Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Lexington, Massachusetts

AD0694552

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

PROBABILITY DENSITY FUNCTION ESTIMATION
(WITH APPLICATIONS TO RECEIVER DESIGN
FOR RECEPTION IN NON-GAUSSIAN NOISE)

J. E. EVANS

Group 66

TECHNICAL NOTE 1969-47

29 AUGUST 1969

LEXINGTON   MASSACHUSETTS

ABSTRACT

This note considers the problem of estimation of an unknown probability density function, p(x), and the derivative of the logarithm of that density function,

$$f(x) = \left[\frac{d}{dx} \ln p(x) \right] ,$$

given N samples from an ensemble whose probability density is p(x). Our principle objective is to estimate f(x) since it has been shown that the optimal receiver for known threshold signals in additive white (but possibly non-Gaussian) noise consists of a filter matched to the signal preceded by a no-memory device whose transfer characteristic is

$$z = f(x) \Big|_{x=y}$$

where y is the input to the device and z the output.

Although our approach is primarily motivated toward obtaining a good estimate of f(x), the method and results would appear also to be applicable to finding "good" estimates of p(x). The "goodness" of the estimation procedure is investigated theoretically and experimentally.

I. INTRODUCTION

This note considers the problem of estimation of an unknown probability density function, p(x), and the derivative of the logarithm of the density function

$$f(x) = \frac{d}{dx} [\ln p(x)] \qquad (1.1)$$

given n samples from an ensemble whose probability density function is p(x). Our principle objective is to find a "good" estimate f(x) because of its use in communications theory applications. For example, it has been shown that the optimal receiver for known threshold signals in white (but possibly non-Gaussian) noise is a filter matched to the signal preceded by a no-memory device whose transfer characteristic is

$$z = f(x) \Big|_{x=y} \qquad (1.2)$$

where y is the input to the device and z the output. Although our approach is primarily motivated toward obtaining a good estimate of f(x), the method and results would appear to be applicable to obtaining "good" estimates of p(x).[*]

The topic of obtaining "good" estimates of the density function has some relevance for our particular problem and so we first review some approaches proposed by other investigators. Next, we present our estimation procedure which is based on the assumption that p(x) is smooth enough over a finite interval ("the smoothing interval") such that $\ln p(x)$ can be accurately represented by a power series with a finite number of terms. This allows us to characterize the problem of finding a "good" estimate of p(x) and f(x) as one of estimating the power series coefficients.

We find that a power series expansion with two terms gives a readily implemented estimation procedure (based on a maximum likelihood estimation)

---

[*] An approach somewhat similar to ours has been suggested by Liden[11] for estimating p(x) in connection with pattern recognition studies.

for estimating $p(x)$ and $f(x)$ as well as lower bounds of the variance of the estimates. The "goodness" of the estimates in terms of bias and variance is examined experimentally using sample functions drawn from known ensembles.

Finally, we indicate possible use of this procedure in communications theory and pattern recognition studies.

II. REVIEW OF PROBABILITY DENSITY FUNCTION ESTIMATION

Despite the importance of problems related to density function estimation, there is comparatively little literature regarding the estimation procedure for the case in which one has only a very vague idea of what the form of the density function might be. At the outset, it must be noted that all smoothing operations of interest assume some regularity of the function being estimated, e.g. the existence of derivatives, or the vanishing of high order Fourier coefficients etc. We somewhat arbitrarily classify prior work we are aware of into three classes:

1.   "Distribution independent" linear smoothing [Bartlett,[3] Parzen,[7] Rosenblatt[8]]: These papers consider the class of estimates of $p(x)$ of the form

$$\hat{p}_n(x) = \int_{-u(n)/2}^{+u(n)/2} \omega_n(x-s)\, d\, P_n(s) \qquad (2.1)$$

where $P_n(x)$ is the sample distribution function for n samples. If $\omega_n(s)$ is taken to be suitably well behaved* even function of s such that

$$\int_{-u(n)/2}^{u(n)/2} \omega_n(s)\, ds = 1$$

it can be shown that by having the width, $u(n)$, of the smoothing interval go to zero as $n \to \infty$ such that $n\, u(n) \to \infty$, then $\hat{p}_n(x)$ converges to $p(x)$ in quadratic mean for all points of continuity of $p(x)$ with the variance of $\hat{p}_n(x)$ going to zero as $\left[n\, u(n) \right]^{-1}$.

---

*Parzen[7] shows that it is sufficient to require that $\omega_n(s) = K\dfrac{s}{u(n)}$ where $K(y)$ meets the conditions:

1. $\sup_{-\infty < y < \infty} |K(y)| < \infty$ 2. $\lim_{y \to \infty} |y\, K(y)| = 0$ 3. $\int_{-\infty}^{\infty} |K(y)|\, dy < \infty$

2.   "Partial distribution dependent" linear smoothing [Whittle[10]]:
The class of estimates of p(x) considered is again of the form

$$\hat{p}_n(x) = \int_{-u(n)/2}^{u(n)/2} \omega_n(x-s) \, d \, P_n(s).$$ (2.2)

However, in this case, the $\omega_n(s)$ are chosen to minimize the variance of $\hat{p}_n(x)$ over a given family of density functions [an example of this would be to obtain the appropriate $\omega_n(s)$ when the apriori p(x) is taken to be Gaussian with unit variance and a mean uniformly distributed between +10 and -10]. For the classes of p(x) considered, it is found that the variance of $\hat{p}_n(x)$ does not go to zero faster than $[n]^{-1}$.

3.   "Distribution dependent" smoothing [Cramer,[4] Parzen,[6] and many others]. In this case, one assumes a model of p(x) that is completely specified except for the values of a finite set of parameters. The values of these parameters are then estimated by methods such as maximum likelihood estimation or the method of moments. In a case where sufficient evidence exists to suggest that a model containing a manageable set of parameters will fit the data, this method probably offers the best procedure. However, in many cases of interest, the number of parameters required and the form of equations that result in obtaining parameter estimates make this method difficult to apply in practice.

The first two approaches are analogous to determination of a suitable linear estimator of a continuous waveform given a received waveform whereas the third approach is analogous to determining the best (e.g., maximum likelihood and/or minimum variance) nonlinear estimator of the continuous waveform. For example, in the case where the apriori density function is taken to be Gaussian distributed with zero mean, the third approach (using maximum likelihood estimation) gives:

$$\hat{p}_n(x) = \left[2\pi \, \hat{V} \right]^{-1/2} \exp \left[-x^2/2 \, \hat{V} \right]$$ (2.3)

where

3

$$\hat{V} = \int_{-\infty}^{\infty} s^2 \, dP_n(s) \quad . \tag{2.4}$$

We note that the estimator of equations 2.3 and 2.4 is not a linear functional of the sample distribution function $P_n(x)$.

For our particular problem, the first two approaches discussed above have the liability that a good estimate of $p(x)$ may not yield particularly good estimates of $f(x) = \dfrac{d \ln p(x)}{dx}$. Furthermore, it is not exactly clear how one would operate on a "good" estimate of $p(x)$ to give a good estimate of $f(x)$. The third approach gives a straightforward procedure for estimating $f(x)$; but as we have indicated, it is often quite difficult to find a manageable form that fits the data over the required range of x.

III. ESTIMATION PROCEDURE

In this section, we present the assumptions made as to the regularity of $p(x)$ and develop in detail an estimation procedure.

The regularity condition that we impose is to consider the logarithm of the density function, $\ln[p(x)]$, to be a reasonably smooth function of x, so that in the vicinity of some value of x, say $x = x_0$, the function $\ln[p(x)]$ can be expanded in a power series in $x - x_0$ with a finite number of power series coefficients. * In this way, estimating the value of $p(x)$ (and parameters such as $\dfrac{d}{dx}[\ln p(x)]$) at x_0 reduces to estimating the power series coefficients for the expansion around x_0.

A model is assumed for the distribution of the sample density function,

---

*After this work was completed I became aware of the work of Liden[11] who proposed the model $\ln p(x) = \sum_{i=1}^{M} \theta_i \varphi_i(x)$ where $\varphi_i(x)$ represent fixed functions and the θ_i are estimated from the data. Liden points out that his model is motivated by a result of Dynkin[12] that if there is a sufficient statistic of finite dimension M that characterizes the density function $p(x)$, then $p(x)$ must be of the form above. However, Liden was not able to solve the resulting estimation equations for any cases that would be appropriate to our problem.

$p_n(x)$, conditioned on the "true" value of this density function being $p(x)$. This allows us to obtain the likelihood ratio for the observed sample density-function distribution conditional on the power series coefficients having certain values. It is then easy to obtain maximum likelihood estimates of the power series coefficients as well as Cramer-Rao bounds on the variance of the estimates.

From the assumption that $\ln[p(x)]$ can be expanded in a power series of finite order M around x_o we have:

$$\ln p(x) = \ln p(x_o) + \frac{d[\ln p(x)]}{dx}\Big|_{x = x_o} (x-x_o) +$$

$$\cdots + \frac{d^{M-1}[\ln p(x)]}{dx^{M-1}}\Big|_{x = x_o} \frac{(x-x_o)^{M-1}}{(M-1)!} \tag{3.1}$$

$$= \sum_{i=0}^{M-1} a_i(x_o)(x-x_o)^i$$

where we have written $a_i(x_o)$ to emphasize that the power series coefficients will, in general, be a function of x_o. Equivalently, we can rewrite (3.1) as:

$$p(x \mid \vec{a}) = \exp\left[\sum_{i=1}^{M-1} a_i(x_o)(x-x_o)^i\right] \tag{3.2}$$

where we have written $p(x \mid \vec{a})$ to indicate that $p(x)$ is a function of the power series coefficients $\{a_i\}$.

Next we consider an appropriate probabilistic model for the distribution of the sample density function $p_n(x)$ conditional on $p(x)$. In the cases of interest to us, the sample density function $p_n(x)$ can be obtained from the relationship.

$$p_n(x) = \frac{P_n(x+\frac{\Delta x}{2}) - P_n(x-\frac{\Delta x}{2})}{\Delta x} \tag{3.3}$$

where the Δx is chosen small enough so that

$$p(x|\vec{a}) \simeq \frac{P(x + \frac{\Delta x}{2}|\vec{a}) - P(x - \frac{\Delta x}{2}|\vec{a})}{\Delta x} \qquad (3.4)$$

where $P(x|\vec{a})$ is the distribution function corresponding to $p(x|\vec{a})$. The quantity $y(x) = n\,p_n(x)\Delta x$ represents the number of samples of $x(t)$ whose values lie in the interval $(x - \frac{\Delta x}{2}, \ x + \frac{\Delta x}{2})$.

Let us define the "smoothing interval" as the closed interval

$$[\,x_o - \frac{u(n)}{2}, \ x_o + \frac{u(n)}{2}\,]$$

along the x-axis in which equation (3.1) holds. Next we divide the smoothing interval into Ns non-overlapping intervals (i.e., "bins") such that $u(n)/Ns$ equals the quantity Δx of equations 3 and 4. If successive values of $x(t)$ are independent, the joint distribution of the Ns values of $y(x)$ corresponding to the various bins in the smoothing interval is then the multinominal distribution with Ns + 1 exclusive events and n trials [Cramer[4](1948)]. The likelihood ratio for the observed sample density function values $p_n(\vec{x})$ in the smoothing interval is then:

$$Pr\,[\,p_n(\vec{x})|\vec{a}\,] = n!\left[\prod_{j=1}^{Ns} \frac{[\Delta x\,p(x_j|\vec{a})]^{y_j}}{y_j!}\right] \qquad (3.5)$$

$$\left\{\frac{1}{(n - \sum\limits_{\ell=1} y_\ell)!}\ \left[1 - \sum_{\ell=1}^{Ns} \Delta x\,p(x_\ell|\vec{a})\right]^{\,n - \sum\limits_{\ell=1}^{Ns} y_\ell}\right\}$$

where $y_m \overset{\Delta}{=} n\,p_n(x_m)\Delta x$.

The maximum likelihood estimates (ML estimates) of the $\vec{a}$ are obtained from the set of simultaneous equations

$$\frac{\partial}{\partial a_i}\left\{P_r\,[\,\vec{p}_n(.)|\vec{a}\,]\right\} = 0 \quad \text{for } i = 0, 1..., M-1 \qquad (3.6)$$

6

Substituting (3.5) into (3.6), we obtain the set of equations

$$\sum_{j=1}^{Ns} \left[\frac{y_j/\Delta x}{p(x_j|\vec{a})} - \frac{n - \sum_{\ell=1}^{Ns} y_\ell}{1 - \sum_{\ell=1}^{Ns} \Delta x\, p(x_\ell|\vec{a})} \right] \Delta x\, \frac{\partial p(x_j|\vec{a})}{\partial a_i} = 0 \qquad (3.7)$$

for $i = 0, 1, 2, \ldots M-1$.

Equation (3.7) cannot be conveniently solved for cases of interest. However, for the usual case of n large and

$$\sum_{\ell=1}^{Ns} \Delta x\, p(x_\ell|\vec{a}) << 1$$

we can set*

$$\frac{n - \sum_{\ell=1}^{Ns} y_\ell}{1 - \sum_{\ell=1}^{Ns} \Delta x\, p(x_\ell|\vec{a})} = n \qquad (3.8)$$

so that equation (7) becomes

$$n \sum_{j}^{Ns} \left[\frac{P_n(x_j)}{p(x_j|\vec{a})} - 1 \right] \Delta x\, \frac{\partial p(x_j|\vec{a})}{\partial a_i} = 0 \qquad (3.9)$$

For later use in bounding the variance of our estimates of the $\vec{a}$, we note that (under the assumption of equation 8) the elements of <u>Fisher's Information Matrix</u> [Van Trees[9](1968)] are

$$J_{i\ell} = E\left[\frac{\partial \ln p\,[P_n(.\,|\vec{a})]}{\partial a_\ell} \cdot \frac{\partial \ln p\,[P_n(.\,|\vec{a})]}{\partial a_i} \right]$$

$$\qquad (3.10)$$

$$= \sum_{j} (n\Delta x) \frac{\partial p(x_j|\vec{a})}{\partial a_\ell} \frac{\partial p(x_j|\vec{a})}{\partial a_i} \frac{1}{p(x_j|\vec{a})}$$

---

*See Appendix A for a discussion of this assumption.

From equations (3.2) and (3.8), we find that the ML estimates of the power series coefficients at $x = x_0$, are found as the solution to the set of equations

$$\sum_j (x_j - x_0)^i e^{\sum_{r=0}^{M-1} \hat{a}_r (x_j - x_0)^r} = \sum_j (x_j - x_0)^i p_n(x_j) \tag{3.11}$$

where $i = 0, 1, 2, \ldots, M-1$. If M is allowed to be greater than 2, this set of equations cannot be solved conveniently. Thus we limit the investigation to the case $M = 2$. This restraint on M then sets a limit on the $\{x_j\}$ to be considered in obtaining the estimate of the $\{a_i\}$ at $x = x_0$: we consider only those $\{x_j\}$ which lie sufficiently close to x_0 such that the density function can be adequately described by the power series expansion of (3.2) with $M = 2$.

In a later section, we indicate the results of some experiments as to what a reasonable maximum range of $\{x_j\}$ around x_0, i.e., size of u(n), is for some well-known density functions. In cases where n is very large, one might well choose a u(n) that depends on n.

In Appendix B, we show that if the density function p(x) and its first two derivatives are continuous and if we shrink the smoothing interval width, u(n), down as $n^{-\alpha_1}$ where $0 < \alpha_1 < 1/3$, then the estimates of p(x) and $f(x) = d/dx \left[\ell n \, p(x) \right]$ using $M = 2$ will be unbiased with variances that $\to 0$ as $n \to \infty$.

For the case $M = 2$, it is useful to define $p_0 = e^{a_0}$, so that $p(\vec{x} \,|\, \vec{a})$ becomes

$$p(x_j \,|\, p_0, a_1) = p_0 \, e^{a_1(x_j - x_0)} \tag{3.12}$$

and $\hat{p}_0$, the ML estimate of p_0, can be interpreted as the ML estimate of p(x) at $x = x_0$. Substituting (3.12) into (3.11) and carrying out the algebra, we obtain

8

$$\hat{a}_1 = g^{-1}\left[\frac{\sum\limits_j (x_j - x_o) P_n(x_j)}{\sum\limits_j P_n(x_j)}\right] \qquad (3.13)$$

$$\hat{p}_o = \frac{\sum\limits_j P_n(x_j)}{\sum\limits_j e^{a_1(x_j - x_o)}} \qquad (3.14)$$

where the function $g(.)$ is defined by the relationship

$$g(\alpha) = \frac{\sum\limits_j (x_j - x_o) e^{\alpha(x_j - x_o)}}{\sum\limits_j e^{\alpha(x_j - x_o)}} \qquad (3.15)$$

and $\hat{a}_1$ is the ML estimate of $f(x)$ at $x = x_o$. The function of $g(\alpha)$ of equation (3.15) must be obtained numerically, but is a function only of the $\{x_j - x_o\}$ and not x_o alone. Thus, if for all x_o, we keep the set $\{x_j - x_o\}$ constant, $g(\alpha)$ and its inverse need to be computed only once as we compute $\hat{a}_o$ and $\hat{a}_1$ for all x_o.

In Fig. 1 we show $g(\alpha)$ for $u(n) = 1.0$, i.e., $\{x_j\}$ covering the x-axis interval $(x_o - .5, x_o + .5)$. This particular $u(n)$ has no particular significance until the relation of x to the statistics of $x(t)$ is known. One x-axis scale for which Fig. 1 is particularly relevant for has X = value of $x(t)$ in root-mean square units from the mean; i.e., $X = [x(t) - \overline{x(t)}]/\sigma_{x(t)}$.

IV. BOUNDS ON THE ESTIMATION ERRORS

Next we consider the Cramer-Rao bounds on the variance of $\hat{p}_o$ and $\hat{a}_1$ assuming that $\hat{p}_o$ and $\hat{a}_1$ are unbiased.[*] Using equations (3.2) and

---

[*]Appendix B shows that by having $u(n)$ and $\Delta x \to 0$ in an appropriate fashion as $n \to \infty$, that $\hat{p}_o$ and $\hat{a}_1$ are asymptotically unbiased. For cases where n is large, but finite, it is difficult to analytically establish whether or not $\hat{p}_o$ and $\hat{a}_1$ are unbiased and/or the degree to which the Cramer-Rao bounds are satisfied with equality in equations (4.1) and (4.3). Thus, we argue that the utility of these bounds can best be established by a consideration of how well they predict the observed deviations for the sample functions drawn from known density functions. In a later section we present some simulation results that indicate that the bounds of equations (4.1) and (4.3) with equality are fairly good.

(3.10), and the well known [4,9] expression for the Cramer-Rao bound with two estimated parameters, we obtain:

$$\text{Var}(p_o) \geq \frac{J_{11}}{J_{00}J_{11} - J_{01}J_{10}} \tag{4.1a}$$

$$= \frac{p_o}{(n\Delta x)\left[\sum_j e^{a_1(x_j-x_o)}\right]\left\{1 - \dfrac{\left[\sum_j (x_j-x_o)e^{a_1(x_j-x_o)}\right]^2}{\left[\sum_j e^{a_1(x_j-x_o)}\sum_j (x_j-x_o)^2 e^{a_1(x_j-x_o)}\right]}\right\}} \tag{4.1b}$$

$$= \frac{(p_o)^2}{\left\{n\sum_j p_o\, e^{a_1(x_j-x_o)}\Delta x\right\}\left\{1 - \dfrac{\left[\sum_j (x_j-x_o)e^{a_1(x_j-x_o)}\right]^2}{\left[\sum_j (x_j-x_o)^2 e^{a_1(x_j-x_o)}\sum_j e^{a_1(x_j-x_o)}\right]}\right\}} \tag{4.1c}$$

where the J_{ik} were defined by equation (3.10). It should be noted that the first term in the denominator of equation (4.1c) is the expected number of samples of x(t) whose values lie in the "smoothing interval", i.e., the x axis interval spanned by the $\{x_j\}$. Because of the normalization implicit in the ML estimate of p_o (eq. 3.14), this first term is numerically equal to the number of samples of x(t) whose values lie in the "smoothing interval". The second factor

$$V_o(a_1) \overset{\Delta}{=} 1 - \frac{\sum_j (x_j-x_o)\, e^{a_1(x_j-x_o)}}{\left[\sum_j (x_j-x_o)^2 e^{a_1(x_j-x_o)}\right]\left[\sum_j e^{a_1(x_j-x_o)}\right]} \tag{4.2}$$

in (4.1c) can be interpreted as a multiplicative factor by which the number of samples is reduced because of the slope (i.e., non-zero $\hat{a}_1$) of $\ln[p(x)]$ near $x = x_o$.

Similarly, the Cramer-Rao bound on the variance of $\hat{a}_1$ is given by

$$\text{Var}\,[\hat{a}_1] \geq \frac{J_{00}}{J_{11}J_{\overline{00}}J_{01}J_{10}} \tag{4.3}$$

$$= \frac{1}{\left\{(n\Delta x)\left[p_0 \sum_j e^{a_1(x_j-x_0)}\right]\right\}\left\{\frac{\sum_j (x_j-x_0)^2 e^{a_1(x_j-x_0)}}{\sum_j e^{a_1(x_j-x_0)}} - \left[\frac{\sum_j (x_j-x_0)e^{a_1(x_j-x_0)}}{\sum_j e^{a_1(x_j-x_0)}}\right]^2\right\}}$$

The first bracketed term in Eq. 4.3 is numerically equal to the average number of samples of $x(t)$ in the "smoothing interval", while the second term

$$V_1(a_1) \triangleq \left\{\frac{\sum_j (x_j-x_0)^2 e^{a_1(x_j-x_0)}}{\sum_j e^{a_1(x_j-x_0)}} - \left[\frac{\sum_j (x_j-x_0) e^{a_1(x_j-x_0)}}{\sum_j e^{a_1(x_j-x_0)}}\right]^2\right\} \tag{4.4}$$

can be interpreted as a multiplicative factor by which one multiplies the number of samples in the smoothing interval.

In Figs. 2 and 3, we plot the multiplicative factors depending on a_1, $V_0(a_1)$ and $V_1(a_1)$, that arose in the expressions for the variance of $\hat{p}_0$ and $\hat{a}_1$ respectively for a symetrical "smoothing interval".[*] Again we point out that the Cramer-Rao bounds represent lower bounds so that the usefulness of equations 4.1 - 4.4 is, at this point, inconclusive. In a later section, we will present results that indicate that the bounds established here are meaningful.

V. APPLICATION TO DETERMINISTIC SAMPLE DENSITY FUNCTION

As indicated earlier, the key assumption that $\ln[f(x)]$ can be adequately represented by a power series of low order was primarily motivated by the

---

[*] The smoothing interval width of 0.5 used for figures 2 and 3 might be appropriate in the case where x represents rms units from the mean.

observation that the logarithm of many common density functions is indeed a smooth function and by the fact that this assumption leads to computationally feasible results. Thus we argue that an appropriate criteria for the utility of this approach is how well it does at estimating $\hat{p}_o$ and $\hat{a}_1$ for various density functions.

In Figs. 4 - 7 we show the results of applying the estimation procedure to a (deterministic) sample density function whose form is

$$p_n(x_j) = \frac{1}{\sqrt{2\pi}} \quad \exp\,[-.5\,x_j^2] \tag{5.1}$$

for various sizes of the "smoothing interval". In Figs. 4 and 6, the "x" represents $p(x)$ at various x values while the smooth line represents the m.l.e. $\hat{p}_o(x)$. In Figs. 5 and 7, the light line represents the actual slope of $\ell n[p(x)]$

$$\frac{d}{dx}\,[\,\ell n\,p(x)]\,\big|_{x=x_o} = -x_o \tag{5.2}$$

while the dark line represents $\hat{a}_1(x)$, the ML estimate of the slope of $\ell n\,p(x)$. By examining the differences between the "true" values of the estimated parameter and the ML estimates of these parameters as a function of "smoothing interval" size for various density functions, one can obtain a measure of the bias to be expected using the procedure outlined here.

The results for the Gaussian fit is also of interest because $e^{-x^2/2}$ can be expanded about any point $x=x_o$:

$$\ell n\,e^{-x^2/2} = -\frac{x_o^2}{2} - x_o(x-x_o) - \frac{(x-x_o)^2}{2}\,. \tag{5.3}$$

We note that had we included the $(x-x_o)^2$ term in our power series expansion for $\ell n\,p(x)$, we could estimate the $p(x)\,\big|_{x=x_o}$ and $\frac{d}{dx}\,\ell n\,p(x)\,\big|_{x=x_o}$ exactly for

12

a sample density function of the form of Eq. (5.1). The deviations of the M L estimates from the "true" values of Eqs. (5.1) and (5.2) as shown in Figs. 4-7 thus give an indication of the error one might expect to arise from using a power series expansion of $\ln p(x)$ consisting of only two terms. From equation (5.3), we see that the error gets larger as $|x-x_0|$ increases; the particular smoothing intervals used for Figs. 4-7 were chosen larger than one might use in practice [particularly near the mode of $p(x)$].

VI. APPLICATION TO NOISE SAMPLE FUNCTIONS

Although in Appendix B we establish that the estimates of $p(x)$ and $f(x)$ are asymptotically unbiased and normally distributed with variances given by equations (4.1)-(4.4), it is extremely difficult to analytically establish any useful results in the small or even medium sample cases. Thus, we now present some results of applying the estimation procedure to sample density functions drawn from a variety of noise processes. In each case we show:

1. a plot of the sample density function $p_n(x)$ and the estimate $\hat{p}_0(x)$ of the density function

2. a plot of the estimate $\hat{p}_0(x)$ and the "true", i.e., ensemble density function $p(x)$ together with the Cramer-Rao variance bounds on $\hat{p}_0(x)$. The variance bound curves plotted correspond to $p(x) \pm 3\sqrt{\text{var}[\hat{p}_0(x)]}$ where $\text{var}(\hat{p}_0)$ is given by equation (4.1).

3. a plot of the estimate, $\hat{a}_1(x)$, of $f(x)$ and the "true", i.e., ensemble $f(x)$ together with the Cramer-Rao variance bounds on $\hat{a}_1(x)$. The variance bound curves plotted correspond to $f(x) \pm 3\sqrt{\text{var}[\hat{a}_1(x)]}$ where $\text{var}(\hat{a}_1)$ is given by equation (4.3).

4. a plot of the expected number of samples in the smoothing interval centered at x as a function of x.

5. a plot of the sample exceedance probability, $1 - F_n(x)$, and the ensemble exceedance probability, $1 - F(x)$, to give some information as to how

13

"typical" the particular sample function used was.

Figures 8-12 show the results for $n = 50,000$ and $u(n) = 1.0$ rms units from the mean for a random number simulation of a noise process with a probability density function of the form $p(x) = \lambda e^{-\lambda x}$ for $x \geq 0$. This particular density function is of interest because the density function in any region of $x > 0$, will be fit exactly by our two term power series representation of $\ln[p(x)]$. We note that both the estimates are significantly different from the true values in the region of width 1.0 corresponding to $x = 0$. This arises because the step in $p(x)$ at $x = 0$ makes the regularity assumption invalid. Over the range of x for which our regularity condition does hold, the estimates and "true" values agree to within the variance bounds.

Figures 13-17 show the results for $n = 50,000$ and $u(n) = 1.0$ rms units from the mean for a random number simulation of a Gaussian noise process. In this case, there are no steps or other abrupt changes in the ensemble density function and so our regularity assumption is approximately valid over the entire range of x. In Figs. 13-17, we see that the estimates agree with the "true" values to within the variance bounds.

Figures 18-22 show the results for $n = 50,000$ and $u(n) = 1.0$ rms units from the mean for a random number simulation of a Pareto process with density function $p(x) = 3/x^4$ for $x > 1$. As in the case of the exponential process of Figs. 8-12 we see the effect of the step in $p(x)$ at $x = 1$ is to cause the estimates $\hat{p}_o(x)$ and $\hat{a}_1(x)$ to differ significantly from the "true" values over an interval of width $u(n)$ centered at the discontinuity. There is also some indication in Figs. 18 and 19 that the estimators are slightly biased (in the sense of giving estimates that consistently are slightly outside the variance bounds in one direction). However, this bias (if it indeed exists) would appear to be quite small.

VII. DISCUSSION

In this section, we discuss a minor extension of our estimation procedure as well as some possible uses of the procedure in communications

and pattern recognition problems. In dealing with density functions of low frequency electromagnetic noise (which is known to be non-Gaussian), we have found it helpful to use a small smoothing interval to estimate data near the mode (since the density function changes very rapidly near the mode) and a much larger smoothing interval for estimation on the tail of the density function (since there is a far smaller number of samples per unit of smoothing interval length out on the tails). Since the estimation procedure is done separately for each bin of the density function, it is easy to make the smoothing interval width a function of distance from the mode [although function $g^{-1}(\alpha)$ of equation (3.15) must be recomputed every time the smoothing interval width is changed].

We have already mentioned the utility of $f(x)$ in receiver design. It can be shown (as will be done in a subsequent report) that the function $f(x)$ is also of use in estimation in non-Gaussian noise as well as in obtaining estimates of the error performance improvement that can be obtained using an optimal receiver in non-Gaussian noise. Finally, we note that our estimation procedure might also be of use in pattern recognition studies of the type considered by Liden.[11]

REFERENCES

1.   Algazi, V. and Lerner, R., "Binary Detection in White Non-Gaussian Noise," Unpublished M.I.T. Lincoln Laboratory memoranda (1964).

2.   Antonov, O., "Optimum Detection of Signals in Non-Gaussian Noise," Radio Eng. and Elec. Phys. 12 541-548 (1967).

3.   Barlett, M. S., "Statistical Estimation of Density Functions," Sankayha Series A 245-254 (1963).

4.   Cramer, H. Mathematic Methods of Statistics, Princeton: Princeton Univ. Press (1946).

5.   Gnedenko, B. V., The Theory of Probability, New York: Chelsea (1962).

6.   Parzen, E., On Models for the Probability of Fatique Failure of a Structure, North Atlantic Treaty Organization Advisory Group for Aeronautical Research and Development, Report 245 (1959).

7.   Parzen, E., "On Estimation of Probability Density Function and Mode," The Annals of Mathematical Statistics, 33 520-531 (1962).

8.   Rosenblatt, M. "Remarks on Some Non-Parametric Estimates of a Density Function," Annals of Mathematical Statistics 27 832-837 (1956).

9.   Van Trees, H. L., Detection, Estimation and Modulation Theory Part I New York: Wiley (1968).

10.   Whittle, P., "On the Smoothing of Probability Density Functions," J. Royal Statistical Society Series B 20 334-343 (1958).

11.   Liden, S., Generalized Histogram Approximations to Multivariate Probability Densities with Applications to Pattern Recognition, Engineering Research Center Arizona State Univ. (1968).

12.   Dynkin, C. B., "Necessary and Sufficient Statistics for a Family Probability Distributions " Selected Transactions in Mathematical Statistics and Probability, 1 17-40 (1961).

APPENDIX A

In this appendix, we discuss briefly the rationale behind the assumption that

$$\frac{n - \sum_{\ell=1}^{Ns} y_\ell}{1 - \sum_{\ell=1}^{Ns} \Delta x \, p(x_\ell | \vec{a})} = n \tag{A1}$$

used in obtaining a solvable set of maximum likelihood equations. Our discussion makes heavy use of three results in Cramer[4] namely that

$$E[y_\ell] = n \Delta x \, p(x_\ell | \vec{a}) \tag{A2}$$

$$E\left\{ [y_\ell - E(y_\ell)]^2 \right\} = n \, [\Delta x \, p(x_\ell | \vec{a})] \, [1 - \Delta x \, p(x_\ell | \vec{a})] \tag{A3}$$

and

$$E\left\{ [y_\ell - E(y_\ell)] \, [y_m - E(y_m)] \right\} = -n(\Delta x)^2 \, p(x_\ell | \vec{a}) \, p(x_m | \vec{a})$$

for $\ell \neq m$.

We note that equation (A1) can be written in the form

$$\frac{n(1 - \sum_\ell p_\ell) - \sum_\ell (y_\ell - n p_\ell)}{1 - \sum_\ell p_\ell} \tag{A4}$$

where $p_\ell \overset{\Delta}{=} \Delta x \, p(x_\ell | \vec{a})$.

Equation (A4) clearly reduces to

$$n\left[1 - \frac{1}{n} \frac{\sum (y_\ell - n p_\ell)}{1 - \sum_\ell p_\ell}\right] \ . \tag{A5}$$

Let us examine the statistics of

$$R \triangleq \frac{1}{n} \frac{\sum_\ell (y_\ell - n p_\ell)}{1 - \sum_\ell p_\ell} \tag{A6}$$

From (A2), it is clear that

$$E[R] = 0.$$

Using (A3) and (A4), we have

$$Var[R] = \frac{1}{n^2 (1 - \sum_\ell p_\ell)^2} \left\{ \sum_\ell E[(y_\ell - n p_\ell)^2] \right\}$$

$$+ \sum_{\ell=1}^{Ns} \sum_{\substack{m=1 \\ m \neq \ell}}^{Ns} E[(y_\ell - n p_\ell)(y_m - n p_m)]$$

$$= \frac{1}{n^2 (1 - \sum_\ell p_\ell)^2} \left[\sum_\ell n p_\ell (1 - p_\ell) - \sum_{\ell=1}^{Ns} \sum_{\substack{m=1 \\ m \neq \ell}}^{Ns} n p_\ell p_m \right]$$

$$= \frac{1}{n (1 - \sum_\ell p_\ell)^2} \left[\sum_\ell p_\ell (1 - p_\ell) - \sum_\ell p_\ell (- p_\ell + \sum_{m=1}^{Ns} p_m) \right]$$

$$= \frac{1}{n (1 - \sum_\ell p_\ell)^2} \left[- \sum_\ell p_\ell (\sum_{m=1}^{Ns} p_m) + \sum_\ell p_\ell \right] \ .$$

Thus we have

$$\text{Var}[R] = \frac{\left(\sum_{\ell=1}^{Ns} p_\ell\right)}{n\left(1 - \sum_\ell p_\ell\right)} \tag{A7}$$

$$= \frac{\text{(Probability of x(t) in smoothing interval)}}{\text{(number of samples)(probability of x(t) not in the smoothing interval)}} \cdot$$

Typically, the probability of x(t) in the smoothing interval is less than 0.2, so that the standard deviation of R is

$$\sigma_R \leq \frac{.05}{\sqrt{n}} \quad . \tag{A8}$$

For cases where the procedure outlined here might be used, we would expect n to be at least 100, for which

$$\sigma_R \leq .05 \quad . \tag{A9}$$

Thus, setting

$$n(1 + R) = n \tag{A10}$$

should be a reasonable approximation for a "worst case" of interest, and represents an approximation that improves as we increase n.

APPENDIX B

Asymptotic Behavior of $\hat{p}_o$ and $\hat{a}_1$

In this appendix, we show that if:

1. the ensemble density function $p(x)$ and its first two derivatives are continuous and finite

2. the width, $u(n)$, of the smoothing interval goes to zero proportional to $n^{-\alpha_1}$ as n (the number of independent samples of x) goes to ∞ (where $0 < \alpha_1 < \frac{1}{3}$).

We can choose the bin width, Δx, so that the estimates of $p(x)$ and $f(x) = \frac{d}{dx}[\ln p(x)]$ will be asymptotically unbiased with

1. the variance of the estimate of $p(x)$ going to zero as $(n^{1-\alpha_1})^{-1}$.

2. the variance of the estimate of $\frac{d}{dx}[\ln p(x)]$ going to zero as $(n^{1-3\alpha_1})^{-1}$.

From the conditions on $p(x)$, we can represent $p(x)$ in the smoothing interval by the power series

$$p(x) = p(x_o) + a_1 p(x_o)(x-x_o) + p''(x_o)\frac{(x-x_o)^2}{2} \tag{B1}$$

where

$$a_1 \triangleq \frac{p'(x_o)}{p(x_o)} = \frac{d}{dx}[\ln p(x)]\big|_{x=x_o} \quad . \tag{B2}$$

The proof goes as follows: we will first show that a random variable closely related to the righthand side of equation 3.13 is asymptotically normal with a mean equal to the true value and a variance that goes to zero as $(n^{1-3\alpha_1})^{-1}$. From this, we will show that the estimate $\hat{a}_1$ is therefore asymptotically normal with a mean value equal to a_1 and a variance that goes

20

to zero as $(n^{1-3\alpha_1})^{-1}$. Finally, we will show that the estimate $\hat{p}_o$ is asymptotically normal with a mean equal to $p_o \overset{\Delta}{=} p(x_o)$ and a variance that goes to zero as $(n^{1-\alpha_1})^{-1}$.

First we consider the statistic

$$R_1 = \left[\frac{12}{u^2(n)}\right] \frac{\displaystyle\sum_{j=1}^{Ns} (x_j - x_o) y_j}{\displaystyle\sum_{j=1}^{Ns} y_j} \tag{B3}$$

where:

$Ns \overset{\Delta}{=}$ number of bins in $u(n) = u(n)/\Delta x$

$y_j \overset{\Delta}{=}$ number of values of $x(t)$ that lie in the bin corresponding to x_j

$x_j = x_o + (i - \frac{Ns}{2}) \Delta x$

$$S_{xy} \overset{\Delta}{=} \frac{12}{nu^3(n)} \sum_{j=1}^{Ns} (x_j - x_o) y_j \tag{B4}$$

$$S_y \overset{\Delta}{=} \frac{1}{nu(n)} \sum_{j=1}^{Ns} y_j \tag{B5}$$

$$\Delta x \overset{\Delta}{=} \text{bin width} = k_2 n^{-\alpha_2} \quad (0 < \alpha_2 < 1) . \tag{B6}$$

Let us define

$$\left. \begin{array}{l} p_j = \Delta x \, p(x_j | \vec{a}) \\[2em] q_j = 1 - p_j \end{array} \right\} \quad j = 1, 2, \ldots, Ns \tag{B7}$$

$$P_L = 1 - \sum_{j=1}^{Ns} p_j = 1 - \int_{x_o - u(n)/2}^{x_o + u(n)/2} p(x | \vec{a}) dx \tag{B8}$$

$$q_L = 1 - P_L \tag{B9}$$

21

It is clear that as $n \to \infty$

$$E(y_j) = np_j = n\Delta x\, p(x_j) \to n\Delta x\, p(x_o) = k_2 n^{1-\alpha 2} p(x_o) \ . \tag{B10}$$

Thus, the $\vec{y}$ meet the conditions of Gnedenko's[5] "local limit theorem" from which we obtain

$$p(\vec{y}\,|\vec{a}) = \left[\prod_{j=1}^{Ns} \frac{e^{-\frac{1}{2}q_j z_j^2}}{\sqrt{2\pi p_j}}\right] e^{-\frac{1}{2}q_L z_L^2} \tag{B11}$$

where $z_j \overset{\Delta}{=} (y_j - np_j)/\sqrt{np_j q_j}$.

$$z_L \overset{\Delta}{=} \left(n - \sum_{j=1}^{Ns} y_j - np_L\right)\Big/\sqrt{np_L q_L} \ .$$

As $n \to \infty$, we have from (B6) through (B10):

$$q_j \to 1$$

$$q_L = u(n)\,(x_o/\vec{a}) \to 0 \ .$$

Thus

$$p(\vec{y}\,|\vec{a}) = \prod_{j=1}^{N} \frac{e^{-\frac{1}{2}\frac{(y_j - np_j)^2}{np_j}}}{\sqrt{2\pi np_j}} \tag{B12}$$

from which we conclude the $\{y_j\}$ are asymptotically normal and independent[*]

---

[*]An interesting side note that indicates the rate at which the y_j become independent is the behavior of correlation coefficient between y_j and y_k ($j \ne k$) as $n \to \infty$

$$\rho_{jk} = \frac{E\{[y_j - E(y_j)]\,[y_k - E(y_k)]\}}{[\mathrm{Var}\,(y_j)\,\mathrm{Var}\,(y_k)]^{1/2}} = -\frac{np_j p_k}{[n^2 p_j q_j p_k q_k]^{1/2}}$$

$$= -\sqrt{\frac{p_j p_k}{q_j q_k}} \to -\Delta x\, p(x_o) = k_2 n^{-\alpha 2} p(x_o)$$

with

$$E(y_j) = np_j \tag{B13}$$

$$\text{Var}(y_j) = np_j . \tag{B14}$$

Thus, the sums S_{xy} and S_y are asymptotically normal with means and variances that are to be calculated.

Let us now compute the mean and variance of S_{xy} and S_y respectively. From the body of the report and Appendix A, we recall that the $\vec{y}$ have a multinomial distribution corresponding to n repetitions and $Ns+1$ exclusive events. We will work out $E(S_{xy})$ in detail to indicate the approach and then state the results for $\text{Var}(S_{xy})$, $E(S_y)$ and $\text{Var}(S_y)$. Let us define $\tilde{x}_j = x_j - x_o$. Then:

$$\frac{nu^3(n)}{12} E(S_{xy}) = \sum_{j=1}^{Ns} \tilde{x}_j E(y_j) = \sum_{j=1}^{Ns} \tilde{x}_j \, n \, p_j$$

$$= \sum_{j=1}^{Ns} \tilde{x}_j [n \, \Delta x \, p(\vec{x}_j | \vec{a})] \tag{B15}$$

$$= \sum_{j=1}^{Ns} \tilde{x}_j \, n \, \Delta x \, [p(x_o) + a_1 p(x_o)(x_j - x_o) + \frac{p''(x_o)}{2}(x_j - x_o)^2]$$

which, as $n \to \infty$, becomes

$$\frac{nu^3(n)}{12} E(S_{xy}) = n \int_{-\frac{u(n)}{2}}^{\frac{u(n)}{2}} \tilde{x} \, [p(x_o) + a_1 p(x_o)(\tilde{x}) + \frac{p''(x_o)}{2}(\tilde{x})^2] \, d\tilde{x}$$

$$= n \left[\frac{(\tilde{x})^2}{2} p(x_o) + \frac{(\tilde{x})^3}{3} a_1 p(x_o) + \frac{(\tilde{x})^4}{4} p''(x_o) \Big|_{-\frac{u(n)}{2}}^{\frac{u(n)}{2}} \right]$$

$$\Rightarrow n \, a_1 p(x_o) \frac{[u(n)]^3}{12} \tag{B16}$$

23

so that

$$E(S_{xy}) = a_1 p(x_o) \quad . \tag{B17}$$

Similarly, in the limit as $u(n) \to 0$, it can be shown that

$$E(S_y) = p(x_o) \tag{B18}$$

$$Var(S_y) = \frac{p(x_o)}{n\, u(n)} \tag{B19}$$

$$Var(S_{xy}) = \frac{12\, p(x_o)}{n\, u^3(n)} \quad . \tag{B20}$$

Substituting $u(n) = k_1 n^{-\alpha_1}$ into (B19) and (B20), we find that as $n \to \infty$,

$$Var(S_y) \to \frac{p(x_o)}{k_1 n^{1-\alpha_1}} \tag{B21}$$

$$Var(S_{xy}) \to \frac{12\, p(x_o)}{k_1^3 n^{1-3\alpha_1}} \quad . \tag{B22}$$

From (B21) and (B22), we conclude that as $n \to \infty$, S_y converges in probability to the constant $p(x_o)$. Thus, from Cramer (pages 254-55)[4], we conclude that the density function of (S_{xy}/S_y) is asymptotically normal with

$$E\left(\frac{S_{xy}}{S_y}\right) = \frac{E(S_{xy})}{p(x_o)} = a_1 \tag{B23}$$

$$Var\left(\frac{S_{xy}}{S_y}\right) = \frac{Var(S_{xy})}{[p(x_o)]^2} = \frac{[12/k_1^3 p(x_o)]}{n^{1-3\alpha_1}} \quad . \tag{B24}$$

Next we wish to consider the statistical behavior of the estimate of a_1:

$$\hat{a}_1 = q^{-1} \left[\frac{u^2(n)}{12} \frac{S_{xy}}{S_y} \right] \tag{B25}$$

where

$$q(\eta) = \frac{\sum_{j=1}^{Ns} (x_j - x_o) e^{\eta(x_j - x_o)}}{\sum_{j=1}^{Ns} e^{\eta(x_j - x_o)}} \qquad \text{(B26)}$$

As $n \to \infty$,

$$q(\eta) = \frac{\sum_{j=1}^{Ns} (x_j - x_o) e^{\eta(x_j - x_o)} \Delta x}{\sum_{j=1}^{Ns} e^{\eta(x_j - x_o)} \Delta x}$$

$$\to \frac{\frac{u(n)}{2} \int_{\frac{u(n)}{2}}^{\frac{u(n)}{2}} \tilde{x} \, e^{\eta \tilde{x}} \, dx}{\int_{-\frac{u(n)}{2}}^{\frac{u(n)}{2}} e^{\eta \tilde{x}} \, d\tilde{x}} \qquad \text{(B27)}$$

Let η_o be the largest (finite) value of η for which we wish to know the value of $q(\eta)$. Then as $n \to \infty$

$$|\eta \tilde{x}| \le |\eta_o| \frac{u(n)}{2} = |\eta_o| (\frac{k_1}{2}) n^{-\alpha_1} \to 0 \qquad \text{(B28)}$$

so that we can make the substitution

$$e^{\eta \tilde{x}} = 1 + \eta \tilde{x} \qquad \text{(B29)}$$

and thus obtain

$$q(\eta) = \frac{u^2(n)}{12} \eta \qquad \text{(B30)}$$

so that for

$$\frac{u^2(n)}{12} \frac{S_{xy}}{S_y} \le \frac{u^2(n)}{12} \eta_o \qquad \text{(B31)}$$

25

we have

$$\hat{a}_1 = (\frac{S_{xy}}{S_y}) \text{ for } \frac{S_{xy}}{S_y} \le \eta_o \quad . \tag{B32}$$

Now from our previous results for the ratio S_{xy}/S_y, we realize that by choosing $\eta_o > a_1$, we can assert that for $0 < \alpha_1 < \frac{1}{3}$,

$$\text{Pr} \left(\frac{S_{xy}}{S_y} \ge \eta_o \right) \Big\} = 0 \tag{B33}$$

n→∞

Thus, $\hat{a}_1$ is asymptotically normal with mean a_1 and variance

$$\sigma_{\hat{a}_1}^2 = \frac{12}{k_1^3 p(x_o) n^{1-3\alpha_1}} \tag{B34}$$

$$= \frac{12}{[\text{expected number of samples of } x(t) \text{ in the smoothing}]^{[u(n)]^2}}$$
$$\text{interval } u(n)$$

Finally, we consider the statistical behavior of our estimate of $p(x_o)$

$$\hat{p}_o = \frac{\frac{1}{n\Delta x} \sum_{j=1}^{Ns} y_j}{\sum_{j=1}^{Ns} e^{\hat{a}_1(x_j-x_o)}} \quad . \tag{B35}$$

As long as $\hat{a}_1$ is less than some finite constant A_1 as n→∞,

$$\sum_{j=1}^{Ns} \Delta x \, e^{\hat{a}_1(x_j-x_o)} \to \int_{-\frac{u(n)}{2}}^{\frac{u(n)}{2}} (1 + \hat{a}_1 x) dx = u(n) \quad . \tag{B36}$$

From our results regarding the asymptotic behavior of $\hat{a}_1$, we know that for any $A_1 > a_1$

26

$$Pr(\hat{a}_1 > A_1) \to 0 \text{ as } n \to \infty.$$

Thus, from (B5), we have

$$\hat{p}_o = \frac{\frac{1}{n}[n\,u(n)\,S_y]}{u(n)} = S_y \tag{B37}$$

so that asymptotically $\hat{p}_o$ is normally distributed with

$$E(\hat{p}_o) = p(x_o) \tag{B38}$$

and

$$Var(\hat{p}_o) = Var[S_y] = \frac{p(x_o)}{n\,u(n)}$$

$$= \frac{p^2(x_o)}{p(x_o)\,n\,u(n)} = \frac{p^2(x_o)}{k_1\,n^{1-\alpha_1}} \tag{B39}$$

$$= \frac{[p(x_o)]^2}{\text{expected number of samples of } x(t) \text{ in the smoothing interval } u(n)}$$

Fig. 1. G(α) versus α.
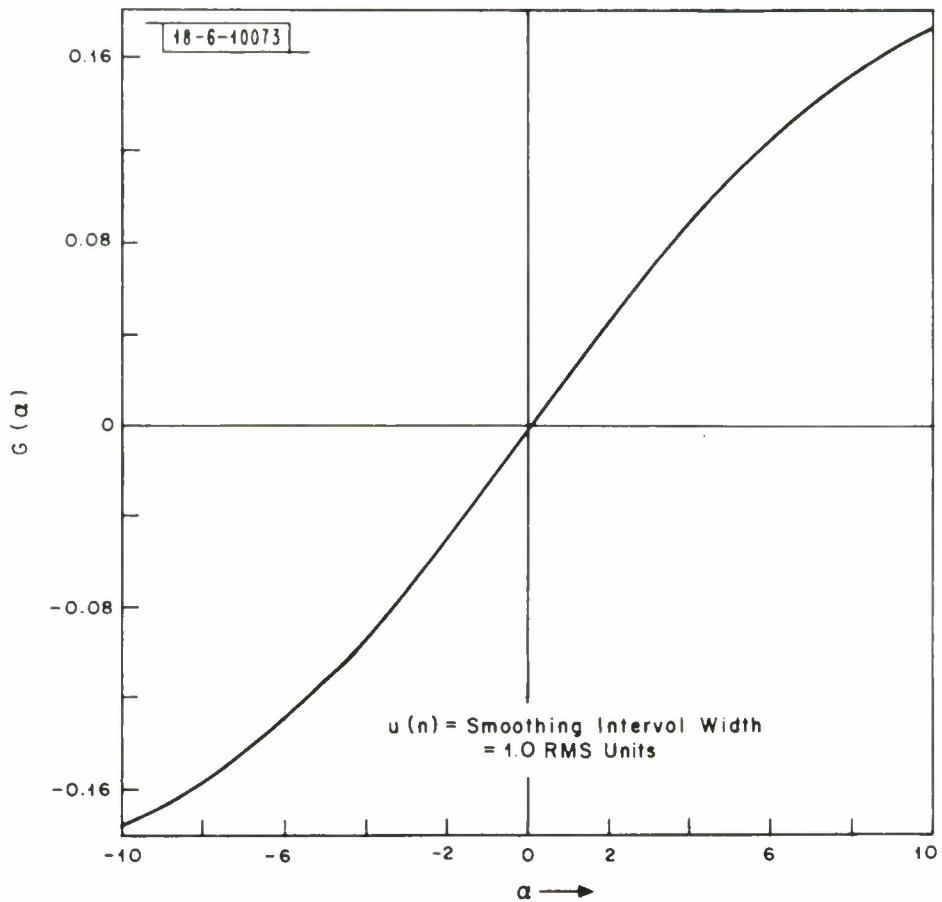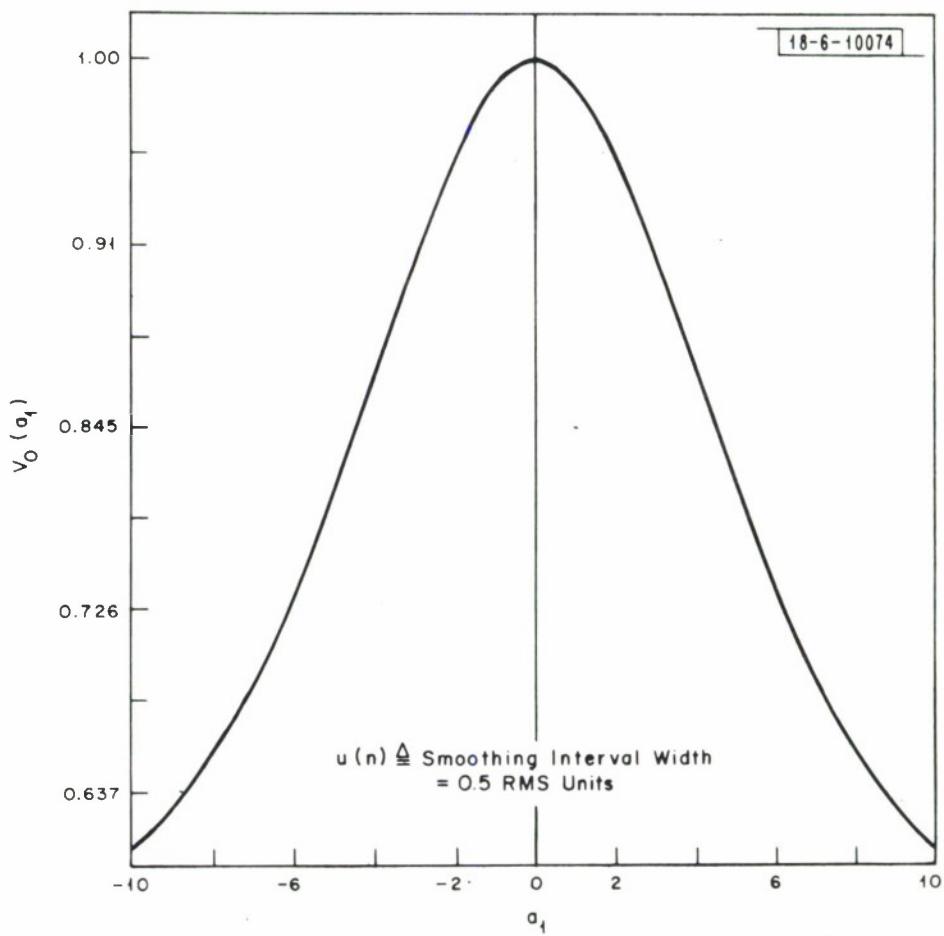
Fig. 2. Density function variance multiplicative factor $V_o(a_1)$ versus a_1.
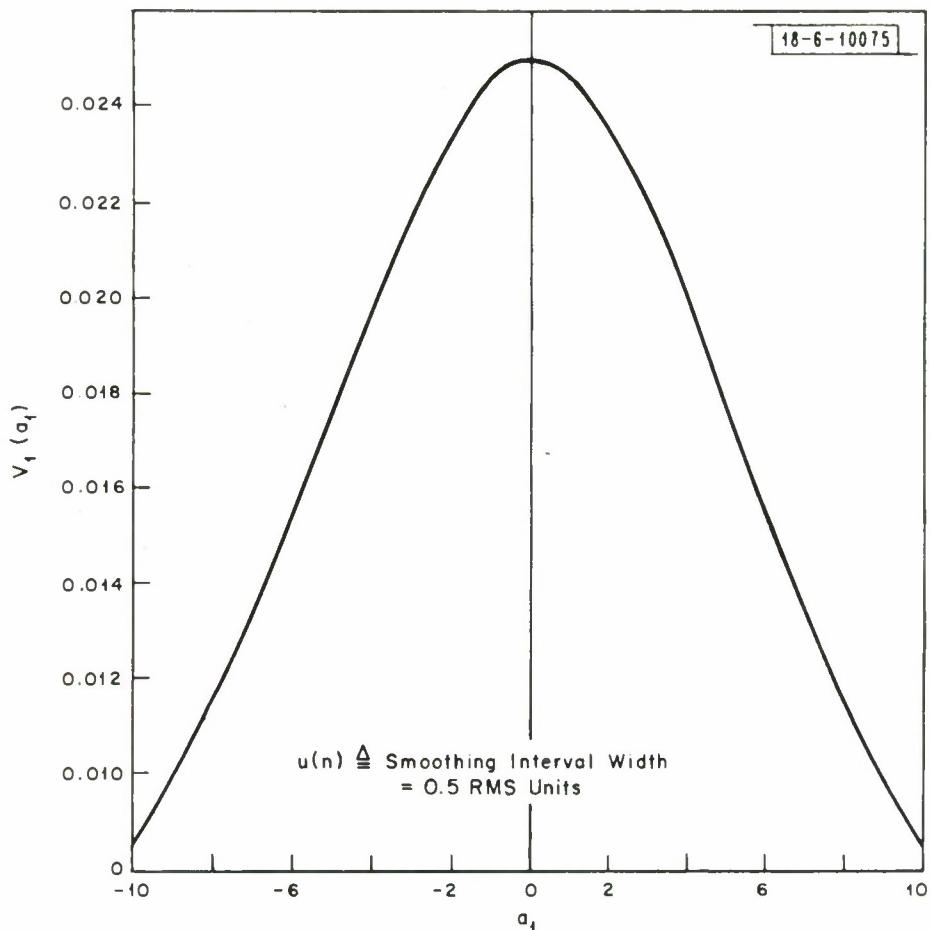
Fig. 3. $\frac{d}{dx}[\ln p(x)]$ variance multiplicative factor $V_1(a_1)$ versus a_1.

30
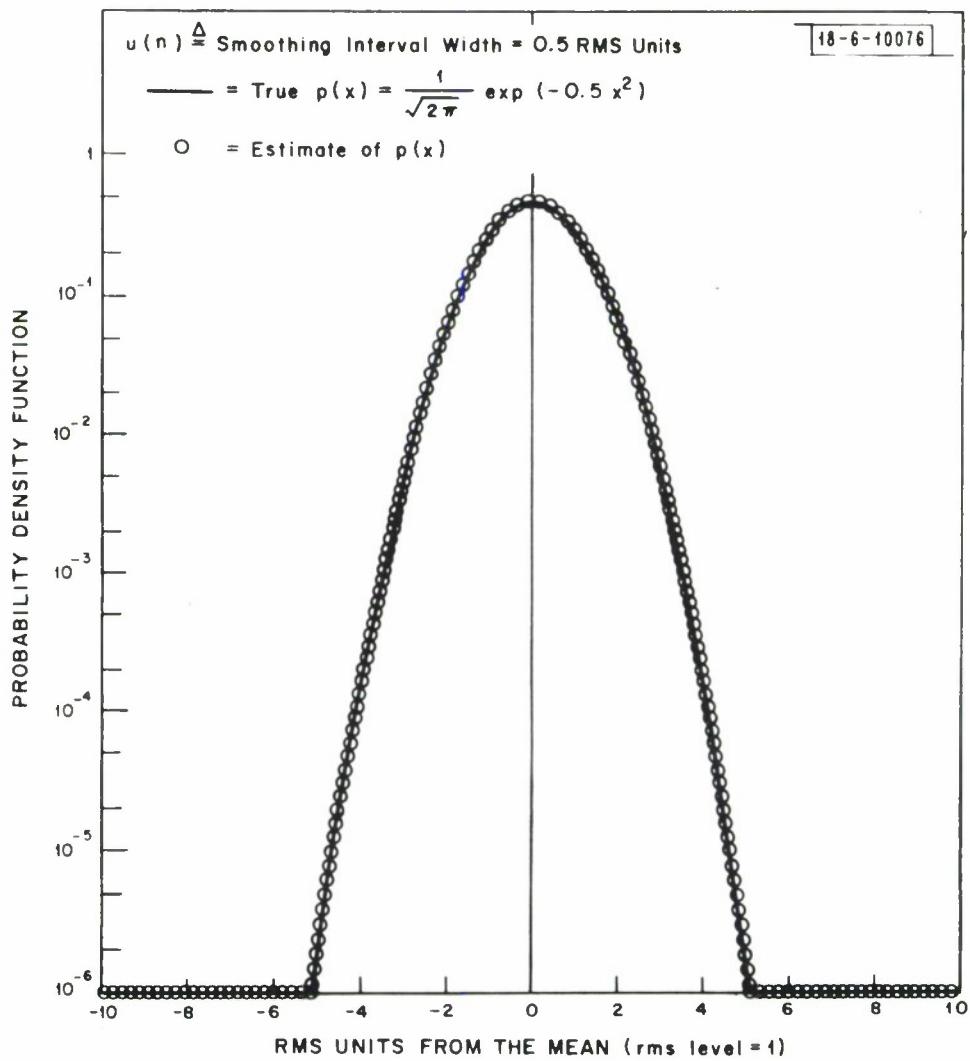
Fig. 4. Results of density function estimation for deterministic sample density function.
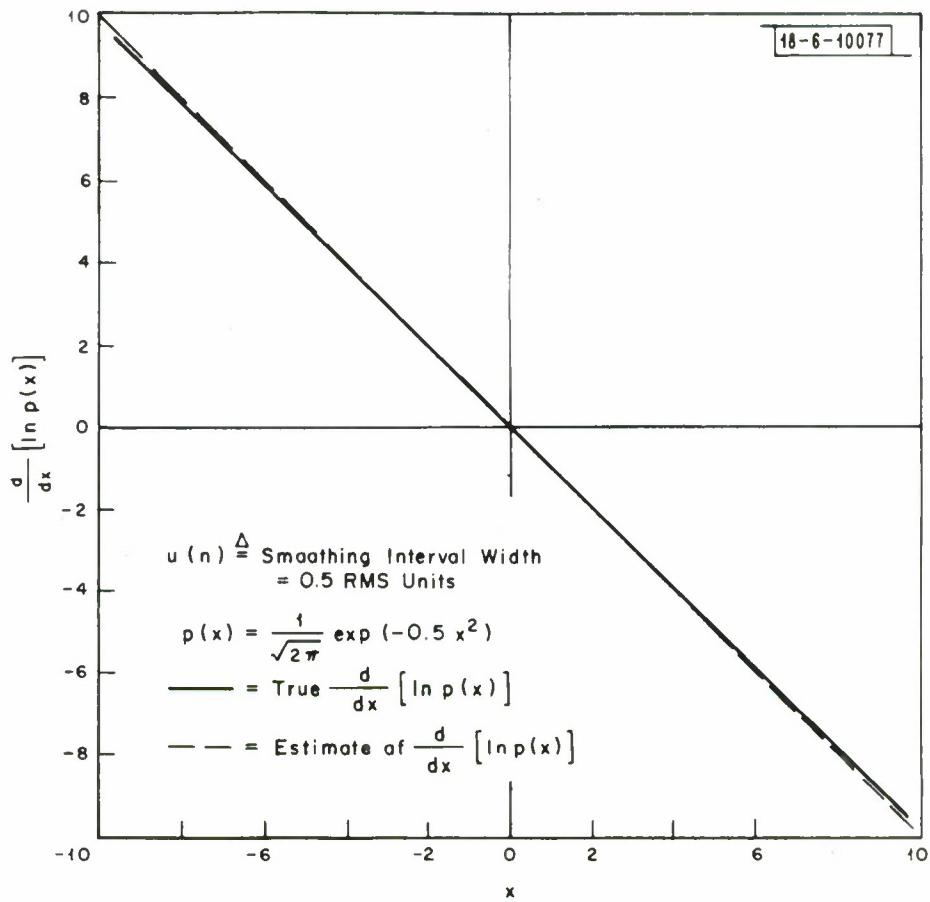
Fig. 5. Results of $\frac{d}{dx}$ [ln p(x)] estimation for deterministic sample density function.
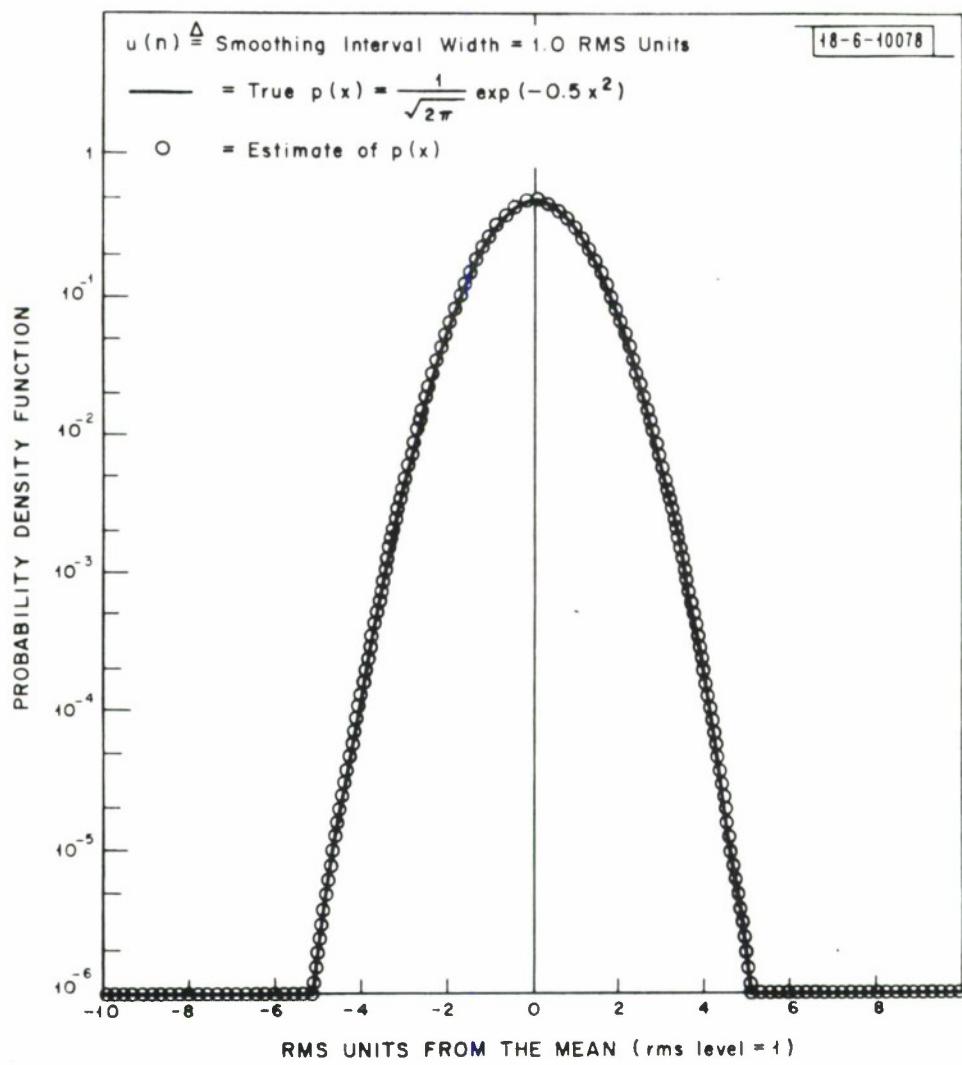
Fig. 6. Results of density function estimation for deterministic sample density function.
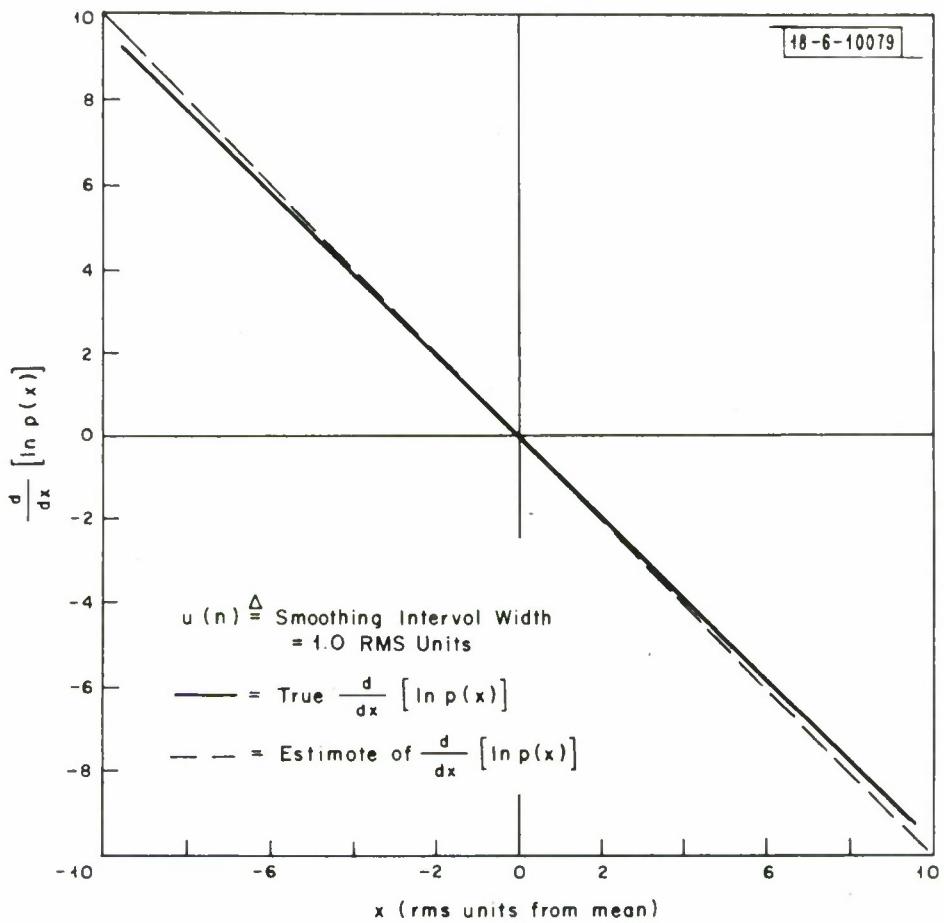
Fig. 7. Results of $\frac{d}{dx}$ [ln p(x)] estimation for deterministic sample density function.

34
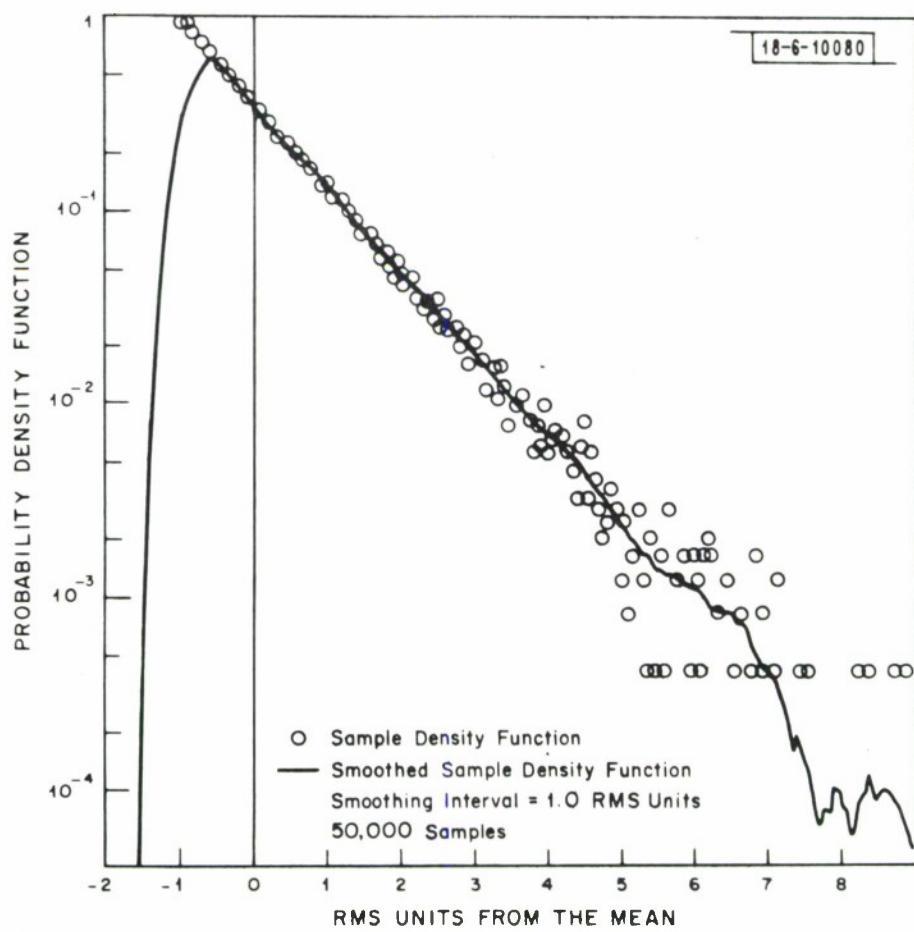
Fig. 8. Sample estimated density functions for
exponential ensemble density function.
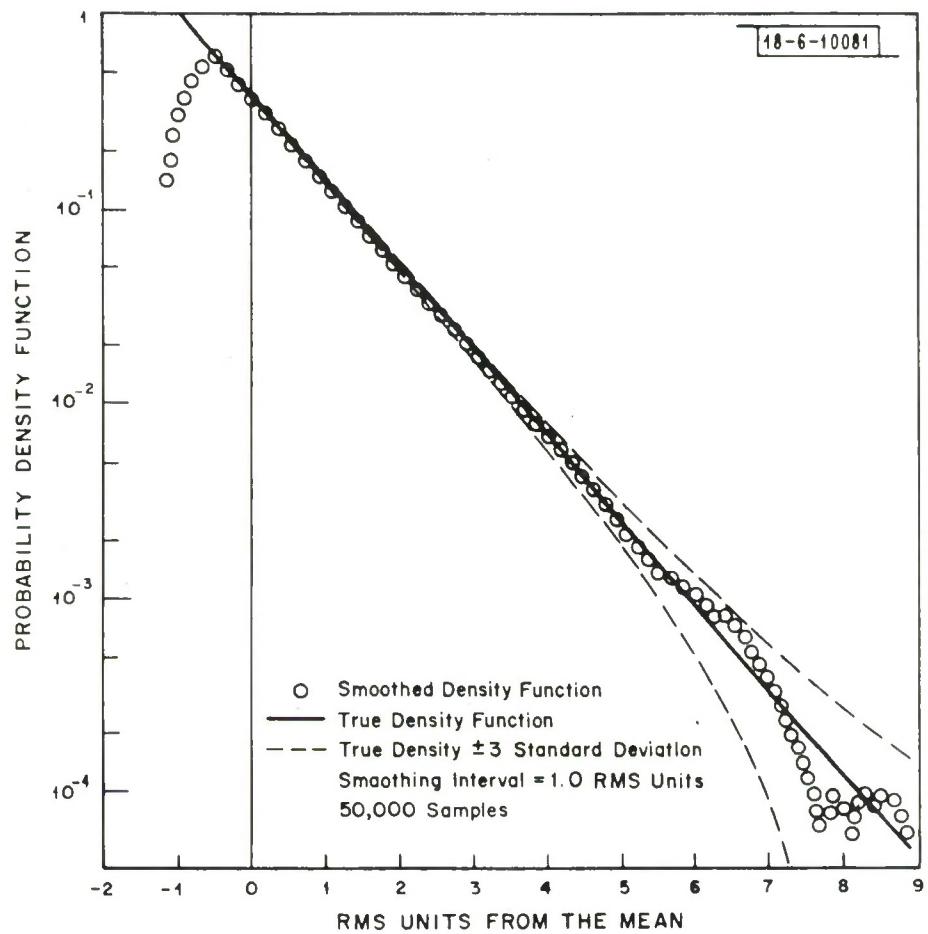
Fig. 9. Comparison of estimated and ensemble density functions
for exponential ensemble density function.
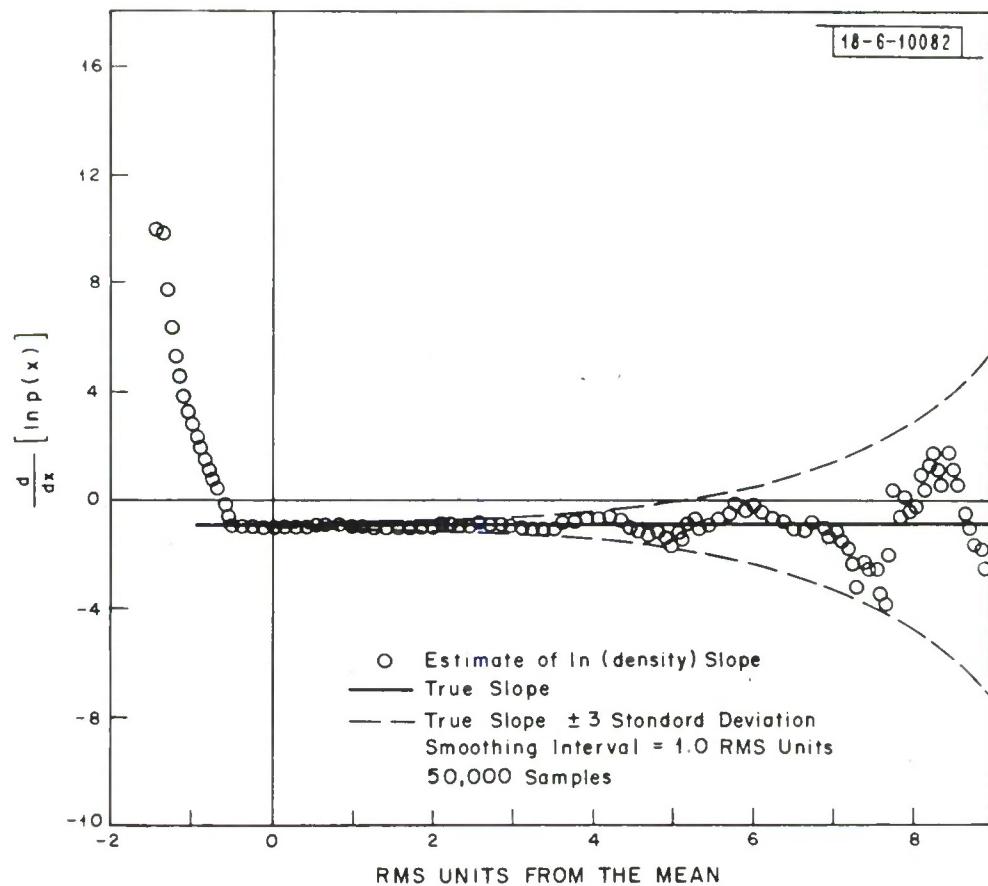
Fig. 10. Comparison of estimated and ensemble $\frac{d}{dx}[\ln p(x)]$ for exponential ensemble density function.
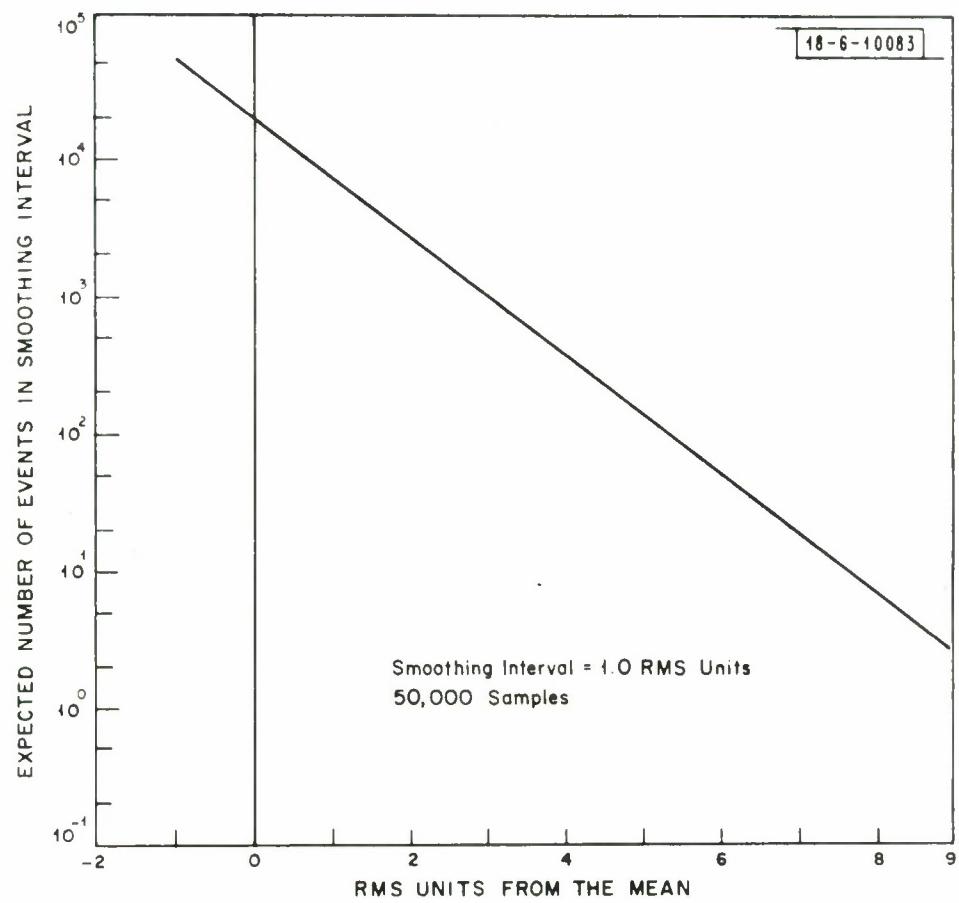
Fig. 11. Expected number of samples in smoothing interval
for exponential ensemble density function.
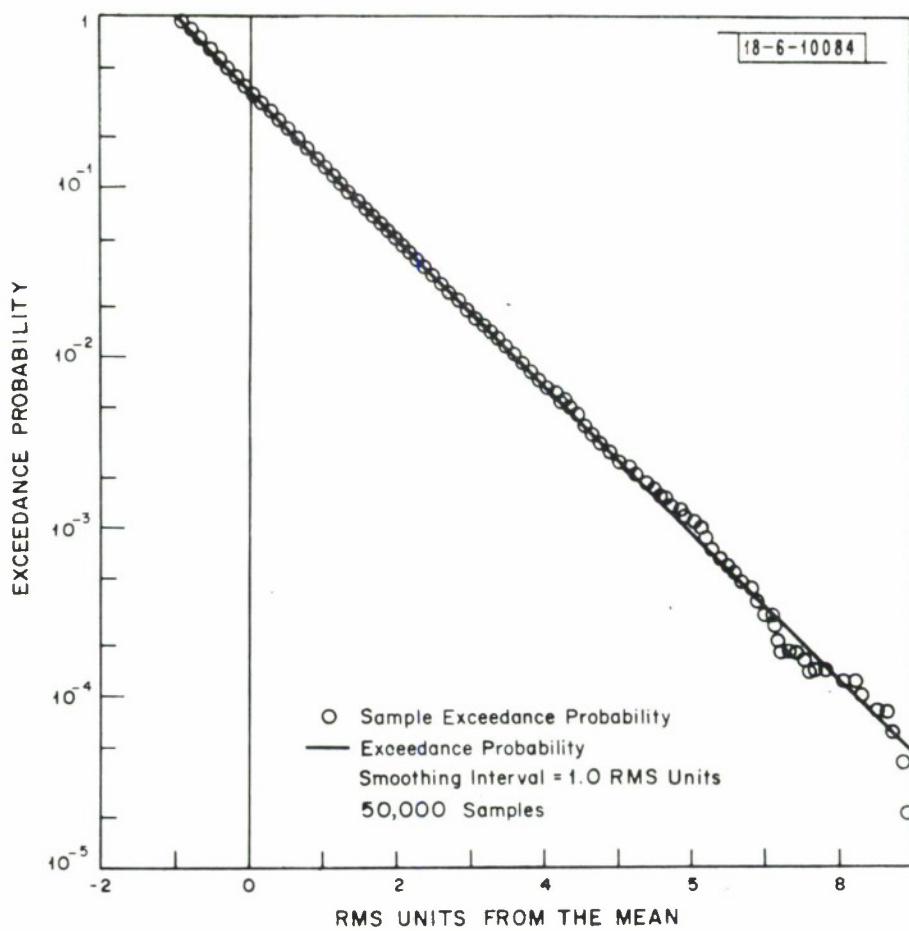
Fig. 12. Comparison of sample and ensemble exceedance probabilities for exponential density function.
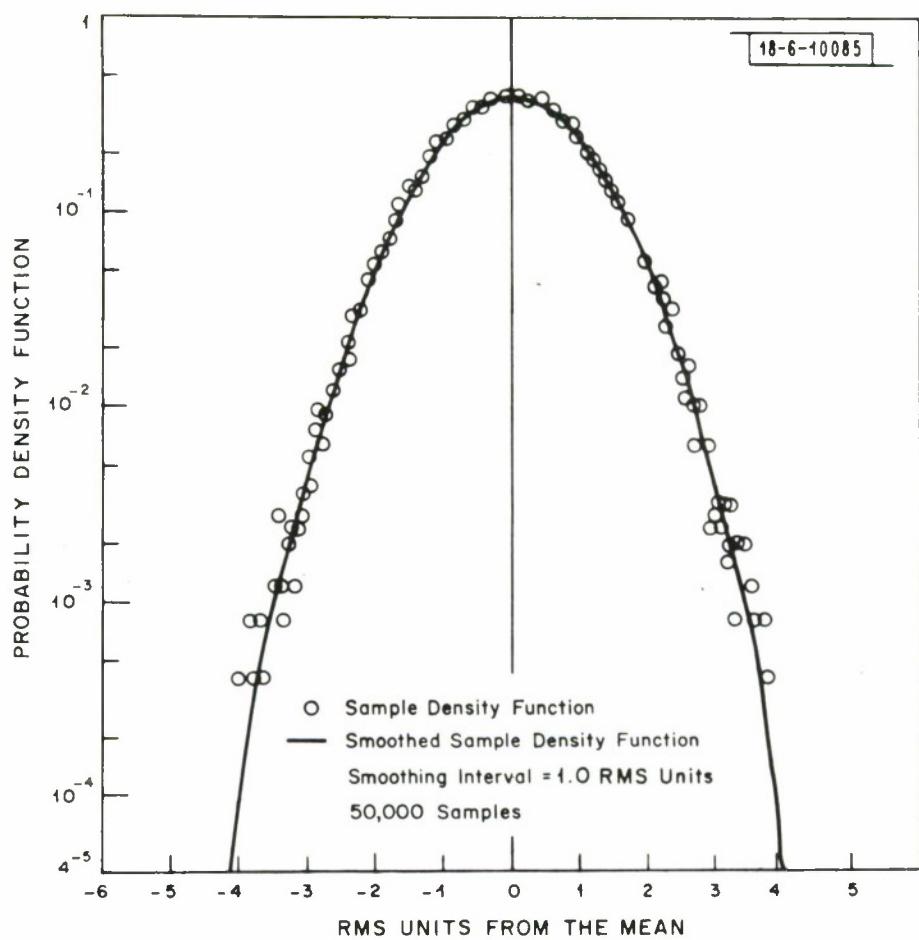
Fig. 13. Sample and estimated density functions for Gaussian ensemble density function.

40
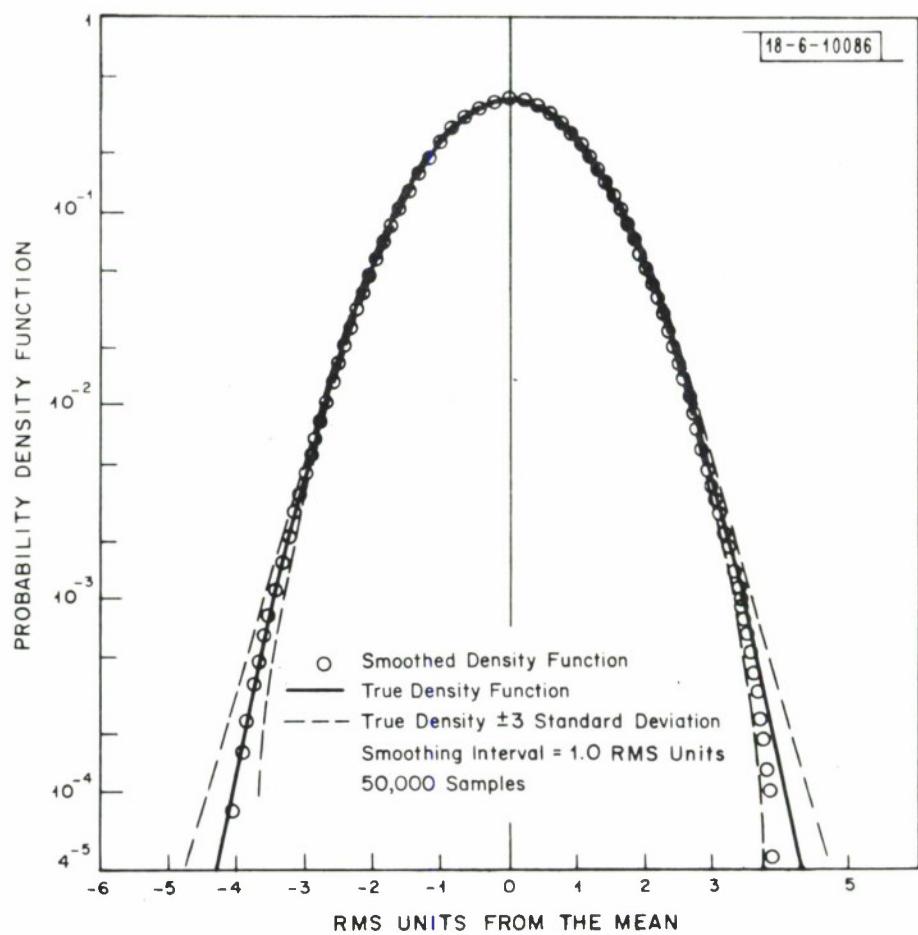
Fig. 14. Comparison of estimated and ensemble density functions for Gaussian ensemble density function.
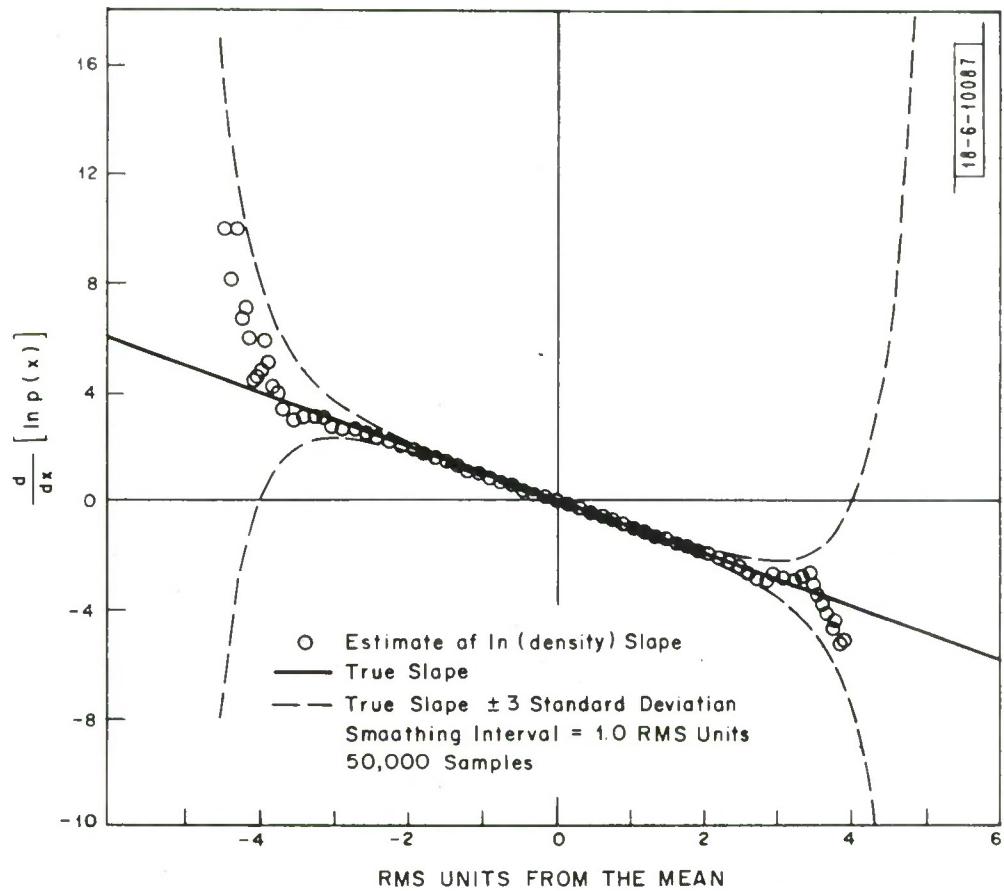
Fig. 15. Comparison of sample and ensemble $\frac{d}{dx}[\ln p(x)]$ for Gaussian ensemble density function.

42
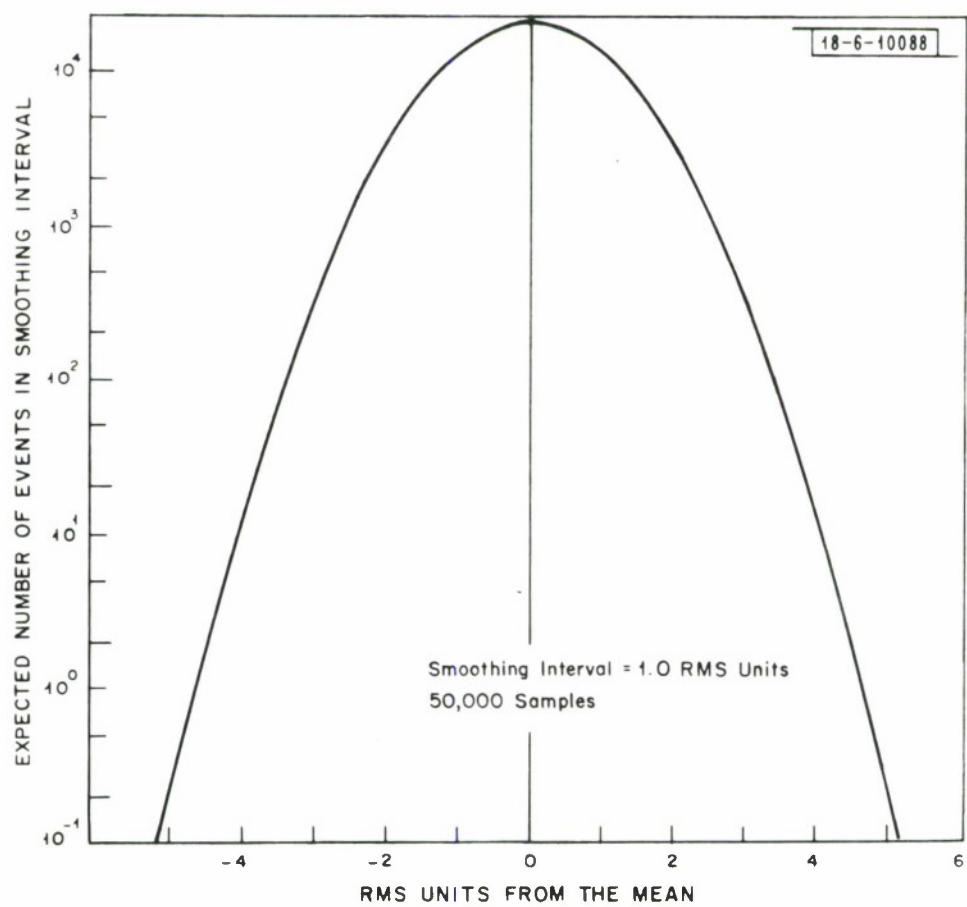
Fig. 16. Expected number of samples in smoothing interval for Gaussian ensemble density function.
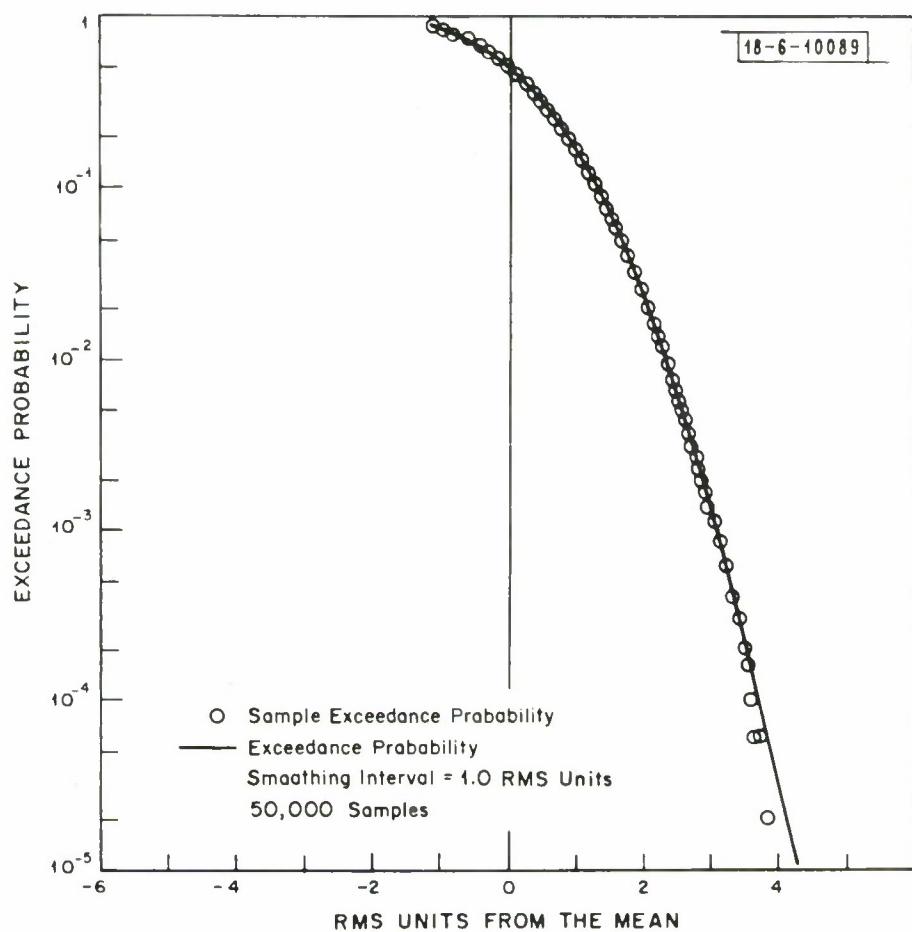
Fig. 17. Comparison of sample and ensemble exceedance probabilities for Gaussian ensemble density function.
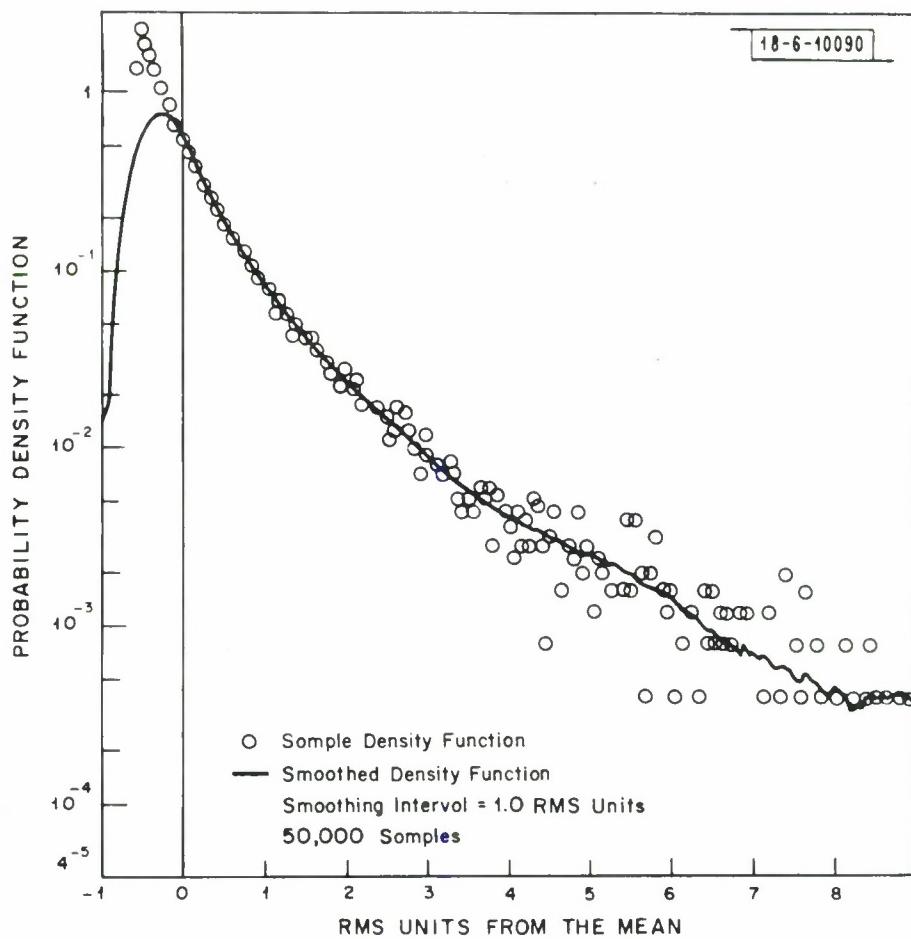
Fig. 18. Sample and estimated density functions for Pareto ensemble density function.
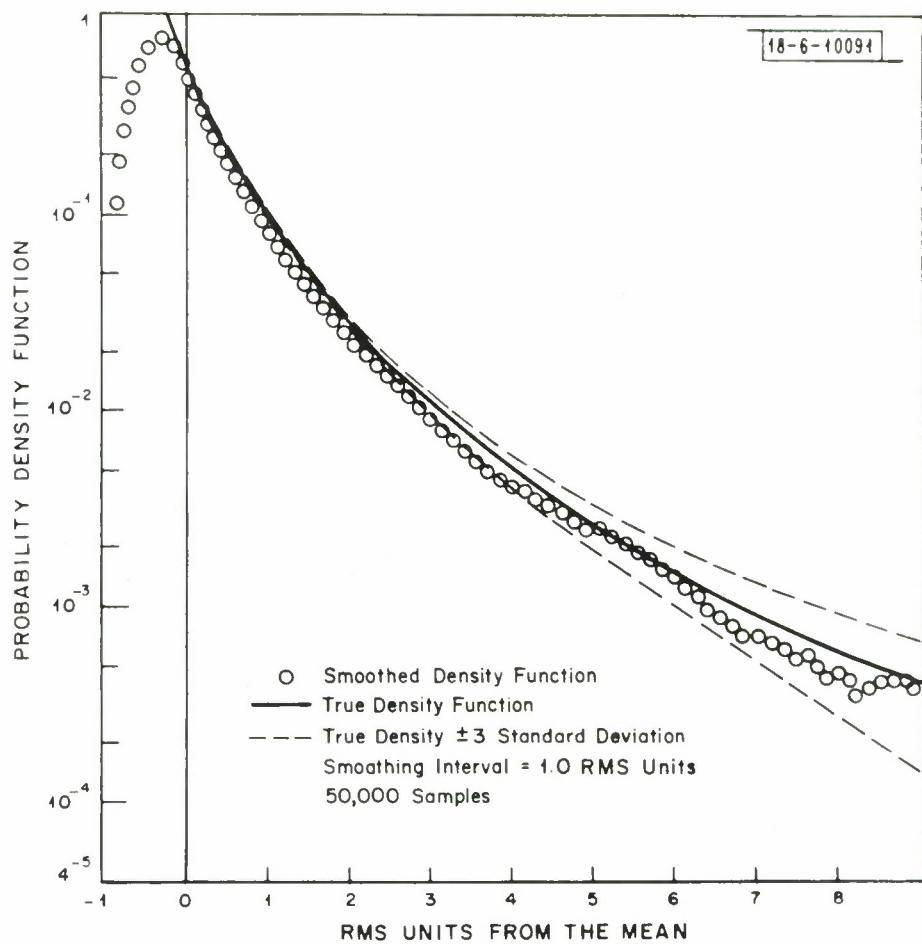
Fig. 19. Comparison of estimated and ensemble density functions for Pareto ensemble density function.
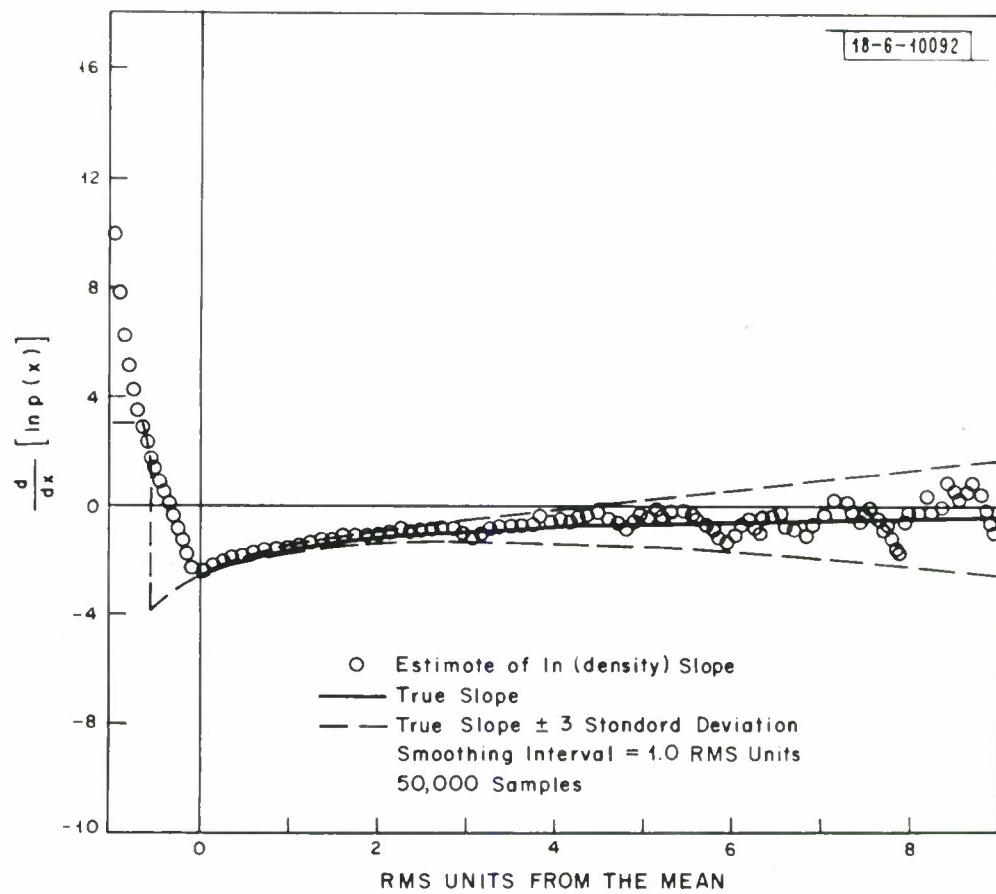
Fig. 20. Comparison of estimated and ensemble $\frac{d}{dx} [\ln p(x)]$ for Pareto ensemble density function.
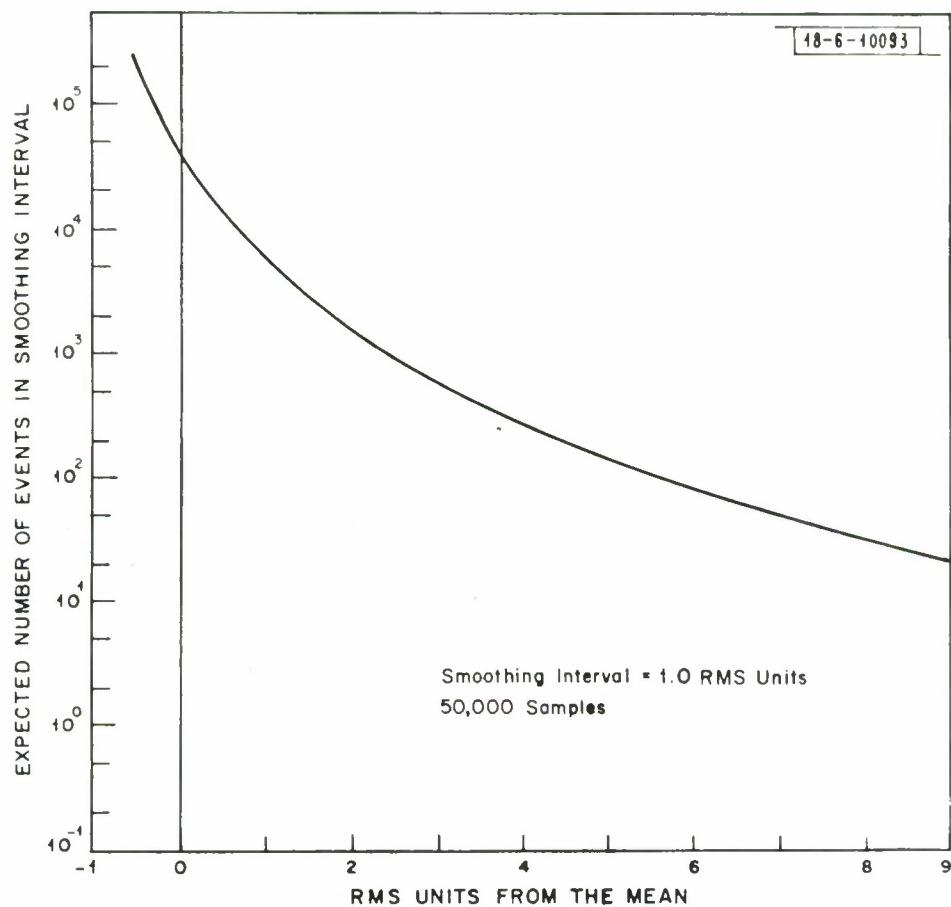
Fig. 21. Expected number of samples in smoothing interval for Pareto ensemble density function.
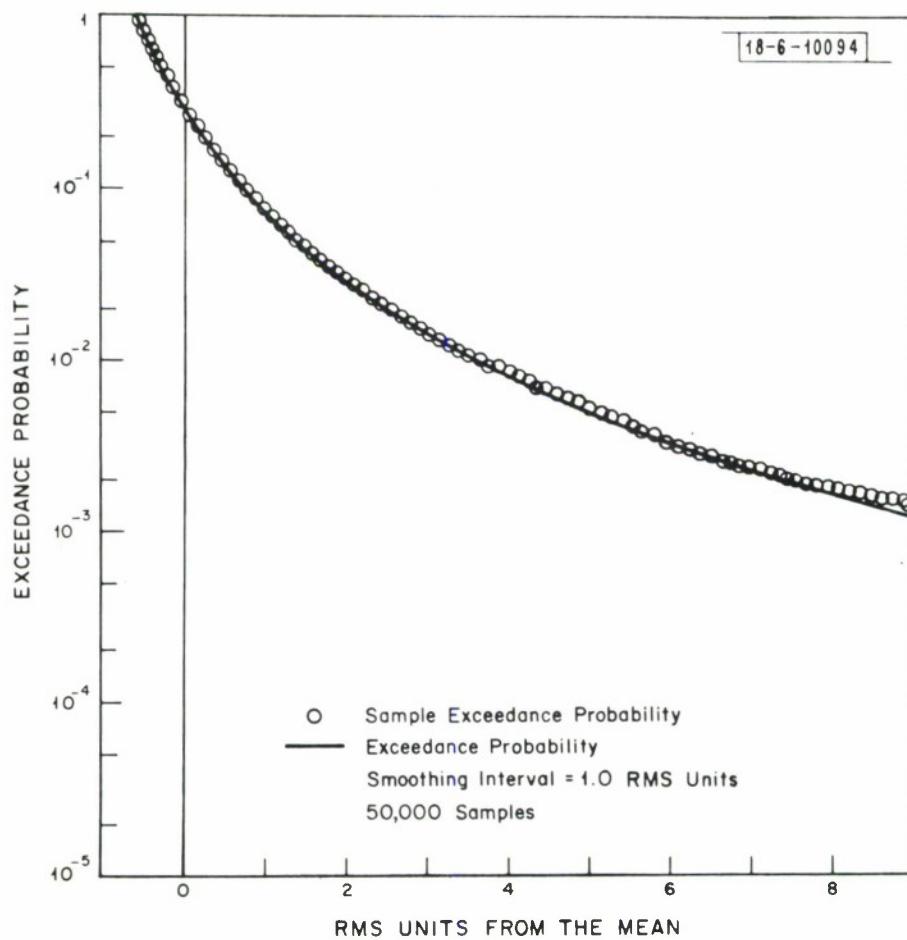
Fig. 22. Comparison of sample and ensemble exceedance probabilities for Pareto ensemble density function.

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Lincoln Laboratory, M.I.T.	Unclassified
	2b. GROUP
	None

3. REPORT TITLE

Probability Density Function Estimation (with Applications to Receiver Design for Reception in Non-Gaussian Noise)

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Technical Note

5. AUTHOR(S) (Last name, first name, initial)

Evans, James E.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
29 August 1969	54	12

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
AF 19(628)-5167	
b. PROJECT NO.	TN 1969-47
1508A	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	ESD-TR-69-243

10. AVAILABILITY/LIMITATION NOTICES

This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
None	Department of the Navy

13. ABSTRACT

This note considers the problem of estimation of an unknown probability density function, $p(x)$, and the derivative of the logarithm of that density function,

$$f(x) = [\frac{d}{dx} \ln p(x)] \quad ,$$

given N samples from an ensemble whose probability density is $p(x)$. Our principle objective is to estimate $f(x)$ since it has been shown that the optimal receiver for known threshold signals in additive white (but possibly non-Gaussian) noise consists of a filter matched to the signal preceded by a no-memory device whose transfer characteristic is

$$z = f(x) \big|_{x=y}$$

where y is the input to the device and z the output.

Although our approach is primarily motivated toward obtaining a good estimate of $f(x)$, the method and results would appear also to be applicable to finding "good" estimates of $p(x)$. The "goodness" of the estimation procedure is investigated theoretically and experimentally.

14. KEY WORDS

probability density function estimation	estimation procedure
non-Gaussian noise	receiver design